

CRANFIELD UNIVERSITY

JOÃO A. DURO

MACHINE LEARNING BASED DECISION SUPPORT FOR A CLASS OF  
MANY-OBJECTIVE OPTIMISATION PROBLEMS

SCHOOL OF APPLIED SCIENCES

PhD Full Time

Supervision:

Dr. Dhish Saxena, Prof. Ashutosh Tiwari, and Dr. Evan Hughes

November 2013

CRANFIELD UNIVERSITY

SCHOOL OF APPLIED SCIENCES

PhD Full Time

JOÃO A. DURO

Machine Learning Based Decision Support for a Class of Many-objective Optimisation  
Problems

Supervision:

Dr. Dhish Saxena, Prof. Ashutosh Tiwari, and Dr. Evan Hughes

November 2013

# Abstract

There is a growing recognition for the multiple criteria decision making (MCDM) based multi-objective evolutionary algorithms (MOEAs) for tackling many-objective optimisation problems (MaOPs). In that, the aim is to utilise the decision makers' (DMs') preferences to guide the *search* towards a few solutions, as against the whole Pareto-optimal front (POF). This thesis is based on the premise that the practical utility of the MCDM based MOEAs may be impaired due to the lack of *objectivity* (a rational basis); *repeatability* (identical preferences for identical options); *consistency* (alike preferences across multiple interaction stages); and *coherence* (alike preferences by multiple DMs) in the DMs' preferences. To counter these limitations, this thesis aimed at developing *offline* and *online* decision support by capturing the preference-structure of the objective functions inherent in the problem model itself. This aim has been realised through the following objectives:

- Identification of a criterion for the decision support: in that, preservation of the correlation-structure of the MOEA solutions is found to be a robust criterion in the case of MaOPs.
- Development of a machine learning based *offline* objective reduction framework: it comprises of linear and nonlinear objective reduction algorithms, which facilitate the decision support through revelation of: (i) the *redundant* objectives (if any), (ii) preference-ranking of the *essential* objectives, (iii) the smallest objective sets corresponding to pre-specified errors, and (iv) the objective sets of pre-specified sizes that correspond to minimum error.
- Development of an *online* objective reduction framework: it addresses a major pitfall associated with the *offline* framework, that—an *essential* objective if erroneously eliminated as redundant, has no scope of being reconsidered in the subsequent analysis. This pitfall is countered through a probabilistic retention of all the objectives, and this serves as a self-correcting mechanism that enhances the overall accuracy.
- *Timing* the decision support: it is acknowledged that the revelations by the decision support may vary depending on when the *offline* or *online* framework is applied during an MOEA run. This uncertainty on the *timing* of the decision support is countered through the proposition of an entropy based *dissimilarity* measure.

The efficacy of the proposed frameworks is investigated against a broad range of test problems (scaled up to 50 objectives) and some real-world MaOPs; and, the accuracy of the corresponding decision support is compared against that of an alternative approach based on preserving the dominance relations. The results illustrate that the proposed frameworks and the corresponding decision support bear significant utility for those MaOPs, where not all the objectives are *essential*, or *equally important* for describing the true POF. The considered real-world problems also bear evidence of the fact that MaOPs with *redundant* and *disparately important* objectives may commonly exist in practice.

# List of Publications

1. Dhish Kumar Saxena, Qingfu Zhang, João A. Duro, and Ashutosh Tiwari. “Framework for Many-objective Test Problems with both Simple and Complicated Pareto-set Shapes”. In *Evolutionary Multi-criterion Optimization*, ser. *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, 2011, vol. 6576, pp. 197-211.
2. Dhish Kumar Saxena, João A. Duro, Ashutosh Tiwari, Kalyanmoy Deb, and Qingfu Zhang. “Objective Reduction in Many-Objective Optimization: Linear and Nonlinear Algorithms,” In *IEEE Transactions on Evolutionary Computation*, vol. 17, no. 1, pp. 77–99, Feb. 2013
3. Dhish Kumar Saxena, Alessandro Rubino, João A. Duro, and Ashutosh Tiwari. “Identifying the Redundant and Ranking the Critical Constraints in Practical Optimization Problems”, In *Engineering Optimization*, vol 45, no. 7, pp. 787–809, Aug. 2012
4. João A. Duro, Dhish Kumar Saxena, Ashutosh Tiwari, Kalyanmoy Deb, and Qingfu Zhang, “Machine Learning based Decision Support for Many-objective Optimization Problems”, submitted in *Neurocomputing*, Elsevier, 2013.



# Acknowledgements

First and foremost, my utmost gratitude to Dr. Dhish Kumar Saxena not only as a supervisor but also as a friend. As his student, I have been extremely lucky to benefit from his very rigorous expertise, constant encouragement, and from his vision. He is a truly remarkable person, and this thesis would not have been possible without his constant support. I am also truly indebted and thankful for all the time that he spent during innumerable technical discussions and also as a father-figure in teaching me how to become a better human-being through: sincerity, hard-work, honesty, ambition, and most important, attitude. I would like also to thank my two co-supervisors, Professor Ashutosh Tiwari and Dr. Evan Hughes, for their constant support, inputs and encouragement during my PhD. I am also grateful to Professor Qingfu Zhang for his expertise and also to Professor Rajkumar Roy for his support. Moreover, I want to thank Dr. Keshav Dahal and Professor Mark Savill for having read and marked this thesis and for their useful detailed comments, positive feedback, and challenging discussion.

I would like to thank all my office colleagues at the Manufacturing Department for their support, friendship and social experience. Many of my colleagues have gone or have just started their work during my PhD, however, I would like to show my gratitude to Dr. Maksim Maksimovic for being a truly remarkable friend and for making my experience a pleasant one. Also, I would like to thank Emanuela Pennetta for being like a mother-figure and for her constant support.

From out of the office I would like to thank Leke Babalola and the rest of the Cranfield

Salsa group together with my Flamenco teacher Eva Rodriguez and my colleague Sara Rivera for making my experience so fun and so enjoyable. I am also obliged to my housemates in Cranfield village namely, Dmitry, Ilya, Alex, David and Susie.

A special thanks goes to my family, my father José Duro, my mother Maria José, and my two brothers Pedro Duro and Felipe Duro. During my PhD my mother won the biggest of all battles against breast cancer and she is now well and back to work. Her recovery gave me inspiration and strength to pursue and complete this thesis. Also, I would like to show my gratitude to both my grandparents Lenine Fialho and Leopoldina Fialho for their support. Finally but not least, I would like to thank my girlfriend Eliza Shaw for being so supportive during my final stage of my PhD.

“In decision-making always ask for a second opinion.”

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Multi-objective Optimisation . . . . .	2
1.1.1	Traditional Approaches . . . . .	4
1.1.2	Evolutionary Approaches . . . . .	6
1.2	Multiple Criteria Decision-Making (MCDM) . . . . .	14
1.3	Many-objective Optimisation: Associated Difficulties . . . . .	16
1.4	MCDM based MOEAs for Many-objective Optimisation: Promise and Pitfalls	19
1.5	Objective Reduction based Decision Support for MCDM based MOEAs . . .	22
1.6	Implementation of the Objective Reduction based Decision Support . . . . .	24
1.7	<i>Timing</i> the Decision Support . . . . .	25
1.8	Thesis Structure . . . . .	27
<b>2</b>	<b>Literature Review</b>	<b>31</b>
2.1	Focus on Evolutionary Many-objective Optimisation . . . . .	34
2.1.1	Assigning Different Ranks to Non-dominated Solutions . . . . .	34
2.1.2	Reduce the Number of Non-dominated Solutions in the Population . . .	37
2.1.3	Using Scalarising Fitness Functions for Fitness Evaluation . . . . .	40
2.1.4	Use of Indicator Functions . . . . .	41
2.1.5	Combining MCDM principles with an MOEA . . . . .	42
2.1.6	Other Methods . . . . .	45

2.2	Many-objective Optimisation Problems: Difficulties Associated with <i>Human</i> Decision-making . . . . .	45
2.3	Objective Reduction based Decision Support . . . . .	48
2.3.1	Objective Reduction Approaches . . . . .	50
2.3.2	Machine Learning based Objective Reduction . . . . .	58
2.3.3	Online Implementation of the Decision Support . . . . .	69
2.3.4	Timing the Decision Support . . . . .	71
2.4	Research Gaps . . . . .	77
2.4.1	Objective Reduction Approaches . . . . .	78
2.4.2	<i>Online</i> Implementation of the Decision Support . . . . .	80
2.4.3	<i>Timing</i> the Decision Support . . . . .	82
<b>3</b>	<b>Research Aim and Objectives</b>	<b>84</b>
3.1	Research Aim and Objectives . . . . .	84
3.2	Methodology . . . . .	86
3.3	Scope of the Research . . . . .	89
3.4	Main Contribution and Key Findings . . . . .	89
3.5	Summary . . . . .	91
<b>4</b>	<b>Machine Learning Based <i>Offline</i> Objective Reduction Framework</b>	<b>93</b>
4.1	Proposed Framework for Linear and Nonlinear Objective Reduction . . . . .	94
4.1.1	Construction of a Positive Semi-definite Matrix . . . . .	96
4.1.2	Eigen-decomposition of the Positive Semi-definite Matrix . . . . .	97
4.1.3	Eigenvalue Analysis . . . . .	97
4.1.4	Reduced Correlation Matrix (RCM) Analysis . . . . .	98
4.1.5	Selection Scheme for Final Reduction based on RCM Analysis . . . . .	100
4.1.6	Computation of Error . . . . .	101
4.2	Generality and Efficiency of the Framework . . . . .	102

4.3	Test Problems . . . . .	104
4.3.1	Redundant Problems . . . . .	105
4.3.2	Non-Redundant Problems . . . . .	106
4.4	Quality Indicators . . . . .	107
4.5	Experimental Settings . . . . .	109
4.6	Feature Selection by the Proposed Algorithms . . . . .	110
4.7	Working of the Proposed Algorithms when Applied to Data Sets Without Noise	112
4.7.1	DTLZ5(3,5): L-PCA and NL-MVU-PCA based on $\mathcal{N}_{\mathcal{P}}$ . . . . .	112
4.7.2	Key Inferences from the Analysis based on $\mathcal{N}_{\mathcal{P}}$ . . . . .	116
4.8	Working of the Proposed Algorithms when Applied to Data Sets Characterised by Noise . . . . .	118
4.8.1	DTLZ5(3,5): L-PCA and NL-MVU-PCA based on $\mathcal{N}_{NS}$ . . . . .	118
4.8.2	DTLZ5(3,5): L-PCA and NL-MVU-PCA based on $\mathcal{N}_{\epsilon}$ . . . . .	122
4.8.3	Discussion of Results . . . . .	126
4.8.4	POF-representation versus POF-approximation . . . . .	128
4.8.5	Benefit of Objective Reduction . . . . .	129
4.9	Experimental Results on a Wide Range of Redundant Test Problems and Comparison with DRP Based Algorithms . . . . .	130
4.10	Experimental Results on a Wide Range of Non-Redundant Test Problems and Comparison with DRP Based Algorithms . . . . .	136
4.11	Customisation of the Proposed Framework . . . . .	145
4.12	Parameter Sensitivity of the Proposed Framework . . . . .	146
4.13	Real World Problems . . . . .	149
4.13.1	Multi-speed Gearbox Design Problem . . . . .	150
4.13.2	Storm Drainage System Problem . . . . .	152
4.14	Comparative Analysis of the Proposed and DRP-based Objective Reduction Algorithms . . . . .	155

4.14.1	Strength and Limitations of the Different Algorithms . . . . .	155
4.14.2	Computational Complexity of Objective Reduction Algorithms . . . .	158
4.14.3	Key Inferences . . . . .	159
4.15	Summary . . . . .	160
<b>5</b>	<b>Decision Support Based on <i>Offline</i> Objective Reduction Framework</b>	<b>164</b>
5.1	Rationale for a Decision Support to Deal with Many-objective Optimisation Problems . . . . .	165
5.2	Proposed Machine Learning Based Framework for Decision Support to the Decision Maker . . . . .	167
5.2.1	Revelation of an <i>Essential</i> Objective Set . . . . .	171
5.2.2	Preference-ranking of all the Objectives . . . . .	171
5.2.3	$\delta$ -MOSS Analysis . . . . .	172
5.2.4	$k$ -EMOSS Analysis . . . . .	172
5.3	Experimental Settings . . . . .	172
5.4	Experimental Results on Radar Waveform Problem . . . . .	174
5.4.1	Demonstration of the Approach-A (NL-MVU-PCA based Framework)	176
5.4.2	Sensitivity of the Decision Support to Different Samples of $\mathcal{N}_{MS}$ . . .	179
5.4.3	Visual Representation of the Results . . . . .	181
5.4.4	Key Highlights of the Decision Support Offered by the Proposed Framework . . . . .	181
5.4.5	Comparative Analysis with the Approach-B (DRP Based Exact Algorithm) . . . . .	182
5.5	Experimental Results on Car Side-Impact Problem . . . . .	185
5.5.1	Demonstration of the Approach-A (NL-MVU-PCA based Framework)	186
5.5.2	Sensitivity of the Decision Support to Different Samples of $\mathcal{N}_{NS}$ . . .	190
5.5.3	Visual Representation of the Results . . . . .	193

5.5.4	Key Highlights of the Decision Support Offered by the Proposed Framework . . . . .	193
5.6	Experimental Results on Storm Drainage System Problem . . . . .	195
5.6.1	Demonstration of the Approach-A (NL-MVU-PCA based Framework)	195
5.6.2	Sensitivity of the Decision Support to Different Samples of $\mathcal{N}_{NS}$ . . .	199
5.6.3	Visual Representation of the Results . . . . .	200
5.6.4	Key Highlights of the Decision Support Offered by the Proposed Framework . . . . .	200
5.7	Experimental Results on Work Roll Cooling Design Problem . . . . .	202
5.7.1	Demonstration of the Approach-A (NL-MVU-PCA based Framework)	203
5.7.2	Sensitivity of the Decision Support to Different Samples of $\mathcal{N}_{NS}$ . . .	207
5.7.3	Visual Representation of the Results . . . . .	209
5.7.4	Key Highlights of the Decision Support Offered by the Proposed Framework . . . . .	209
5.8	Influence of the Used Distance Measure on the Accuracy of the Decision Support	211
5.8.1	Distance Measure in the Context of Approach-A . . . . .	212
5.8.2	Distance Measure in the Context of Approach-B . . . . .	213
5.9	Comparison of the Efficacy of Approach-A and -B in Handling Noise . . . .	216
5.10	Scope of the Decision Support Offered by the Proposed Framework . . . .	222
5.11	Summary . . . . .	223
<b>6</b>	<b><i>Online</i> Objective Reduction Framework</b>	<b>226</b>
6.1	<i>Online</i> Objective Reduction Framework . . . . .	227
6.1.1	Objective Reduction . . . . .	228
6.1.2	Post-processing: Probabilistic Objective Selection Scheme . . . . .	229
6.1.3	<i>Online</i> Implementation with Probabilistic Objective Sets . . . . .	231
6.2	Experimental Settings and Quality Indicators . . . . .	234
6.3	Demonstration of the Framework 4 on DTLZ5(2,5) . . . . .	234

6.3.1	Objective Reduction . . . . .	235
6.3.2	Post-processing: Probabilistic objective selection scheme . . . . .	236
6.3.3	<i>Online</i> Implementation with Probabilistic Objective Sets . . . . .	237
6.4	Handling of Erroneous Preference Information: <i>Online</i> Framework 4 versus <i>Offline</i> Framework 1 . . . . .	239
6.5	Broader Set of Experimental Results . . . . .	242
6.6	Summary . . . . .	244
<b>7</b>	<b><i>Timing</i> the Decision Support</b>	<b>247</b>
7.1	Entropy and Relative Entropy . . . . .	248
7.2	Probability Distribution Estimation . . . . .	249
7.3	Multidimensional Histogram Algorithm for MOEA Populations . . . . .	251
7.3.1	Computationally Efficient Data Structures . . . . .	252
7.3.2	Assignment of a Unique Identification Number to a Cell . . . . .	254
7.3.3	Multidimensional Histogram Algorithm: General Steps . . . . .	255
7.4	Proposed Dissimilarity Measure . . . . .	258
7.5	MOEA-termination Algorithm . . . . .	261
7.6	Test Problems and Experimental Settings . . . . .	265
7.6.1	Test problems . . . . .	265
7.6.2	Experimental Settings . . . . .	265
7.7	Experimental Results . . . . .	270
7.7.1	Experimental Results for Two- and Three-objective Problems . . . . .	270
7.7.2	Experimental Results for Many-objective Test Problems . . . . .	280
7.7.3	Real-World Problems . . . . .	287
7.8	Summary . . . . .	291
<b>8</b>	<b>Discussion and Conclusions</b>	<b>292</b>
8.1	Discussion . . . . .	293



8.1.1	Contribution to Knowledge . . . . .	294
8.1.2	Experimental Validation and Key Findings . . . . .	301
8.2	Conclusions . . . . .	304
8.3	Limitations and Future Work . . . . .	307
8.3.1	Limitations . . . . .	307
8.3.2	Scope for Future Work . . . . .	308

# Nomenclature

symbol	meaning	page
$\sim$	denotes <i>is distributed according to</i>	230
$\triangleright$	denotes <i>more-important-than</i> relation	178
$\chi$	kernel function	65
$c_i^M$	contribution of the $i$ th objective along the principal components	97
$C_X$	covariance matrix	61
$\delta$	error specified for the $\delta$ -MOSS analysis	54
$\delta_{max}$	maximum error required to change the dominance structure	54
$\mathcal{D}$	feasible region in $\Omega$ or simply the search space	2
$\mathcal{D}(p, q)$	<i>dissimilarity</i> measure	259
$\mathcal{D}(p, q)_{\mathcal{I}}$	<i>dissimilarity</i> measure corresponding to $J_I$	258
$\mathcal{D}(p, q)_{\mathcal{Y}}$	<i>dissimilarity</i> measure corresponding to both $P_{NI}$ and $Q_{NI}$	259
$\mathcal{D}(p, q)_{\mathcal{Y}_P}$	<i>dissimilarity</i> measure corresponding to $P_{NI}$	259
$\mathcal{D}(p, q)_{\mathcal{Y}_Q}$	<i>dissimilarity</i> measure corresponding to $Q_{NI}$	259
$\mathcal{D}_{\mathcal{A}}$	denotes the obtained dispersal of solutions	108
$\mathcal{D}_{\mathcal{T}}$	denotes the dispersal of solutions on the true POF	108
$\eta_{ij}$	denotes whether inputs $x_i$ and $x_j$ are $k$ -nearest neighbours	67
$e_i$	normalised eigenvalue corresponding to the $i$ th principal component	97
$\mathcal{E}_i$	error incurred by the $i$ th objective	170

$\mathcal{E}_i^n$	normalised $\mathcal{E}_i$	171
$\mathcal{E}_k^n$	$k$ -minimal error identified by the $k$ -EMOSS analysis	172
$\mathcal{E}_{redn}$	error incurred when $\mathcal{F}_{redn}$ is omitted	171
$\mathcal{E}_t$	error incurred by iteration $t$	102
$\mathcal{E}_T$	total error incurred after $T$ iterations	102
$f_{ij}$	denotes the $i$ th component of the $j$ th principal component	97
$\mathcal{F}_0$	original (or initial) objective set	48
$\mathcal{F}_{\{\delta\}s}$	$\delta$ -minimal set identified by the $\delta$ -MOSS analysis	172
$\mathcal{F}_e$	objective set identified after eigenvalues analysis	98
$\mathcal{F}_{\{k\}s}$	$k$ -minimal set identified by the $k$ -EMOSS analysis	172
$\mathcal{F}_p$	probabilistic objective set	230
$\mathcal{F}_{redn}$	redundant objective set	48
$\mathcal{F}_s$	objective set identified after one iteration of the framework	101
$\mathcal{F}_t$	objective set after iteration $t$	96
$\mathcal{F}_{\mathcal{T}}$	essential objective set	48
$g$	convergence measure for DTLZ problems	108
$\mathcal{H}$	set of objective vectors	231
$\mathcal{H}(X)$	Entropy of the probability distribution $X$	248
$I_s$	normalised maximum spread indicator	108
$J$	number of inequality constraints in a MOP	2
$J_I$	set of cells (intersection region)	251
$\kappa$	number of distance parameters in a test problem	107
$K$	kernel matrix	67
$\mathcal{KL}(p  q)$	Kullback–Leibler divergence between distributions $p$ and $q$	249
$\lambda_i$	eigenvalue corresponding to the $i$ th principal component	97
$L$	number of equality constraints in a MOP	2
$L_1$	Manhattan norm	211

$L_2$	Euclidean norm	211
$m$	number of essential objectives	48
$M$	number of objective functions in a MOP	2
$M_{2\sigma}$	number of eigenvalues required to account for 95.4% of the variance	100
$n$	number of decision variables	2
$n_{bin}$	number of bins per dimension	251
$n_p$	pre-specified number of decimal places	261
$n_s$	pre-specified number of successive generations	261
$N$	number of solutions in a MOEA population	16
$N_g$	number of generations	96
$N_{gmax}$	maximum number of generations	233
$N_{gt}$	number of generations to be determined	248
$N_v$	number of principal components required to account for $\theta$	97
$N_W$	number of generations per iteration	231
$\mathcal{N}_\epsilon$	set of non-dominated solutions generated by $\epsilon$ -MOEA	110
$\mathcal{N}_{\epsilon-p-On}$	set of non-dominated solutions generated by the <i>online</i> framework	234
$\mathcal{N}_{\epsilon-Off}$	set of non-dominated solutions generated by the <i>offline</i> framework	234
$\mathcal{N}_{MS}$	set of non-dominated solutions generated by MSOPS-II	176
$\mathcal{N}_{NS}$	set of non-dominated solutions generated by NSGA-II	48
$\mathcal{N}_{\mathcal{P}}$	set of non-dominated solutions sampled on the true POF	109
$\mathcal{N}_{\mathcal{R}}$	set of non-dominated solutions randomly sampled	110
$\Omega$	decision (variable) space	2
$\mathbb{P}$	denotes the confidence level on $\mathcal{F}_s$	230
$P_{NI}$	set of cells (non-intersection region) for data set $P$	251
$Q_{NI}$	set of cells (non-intersection region) for data set $Q$	251
$\rho$	number of position parameters in a test problem	107
$\mathcal{R}$	random number	230

$R$	correlation matrix	63
$sc_i$	selection score corresponding to the $i$ th objective	101
$\mathcal{S}_i$	denotes the $i$ th subset of <i>identically</i> correlated objectives	99
$\hat{\mathcal{S}}_i$	denotes the $i$ th subset of <i>potentially identically</i> correlated objectives	99
$\theta$	variance threshold	97
$T_{cor}$	correlation threshold	99
$U(0, 1)$	uniform distribution over the interval $[0, 1]$	230
$\varphi$	fraction of the redundant space	230
$V_i$	denotes the $i$ th eigenvector	97
$\mathbf{x}$	vector of decision variables	2
$\mathbf{w}$	vector of preference-weights	40
$\mathcal{W}_{redn}$	collective preference-weight for $\mathcal{F}_{redn}$	230
$\mathcal{W}_s$	collective preference-weight for $\mathcal{F}_s$	230

# Abbreviations

abbreviation	meaning	page
AR	average ranking	34
ARF	average ranking front	34
CTP	constrained test problems	266
$\delta$ -MOSS	$\delta$ -minimum objective subset	54
DFs	desirability functions	44
DM	decision maker	3
DRP	dominance relation preservation	23
DTLZ	Deb-Thiele-Laumanns-Zitzler test problems	45
EMO	evolutionary multi-objective optimisation	6
$\epsilon$ -MOEA	epsilon multi-objective evolutionary algorithm	12
FHI	fitness homogeneity indicator	73
FR	Favour relation	36
$k$ -EMOSS	minimum objective subset of size $k$ with minimum error	54
KKT	Karush–Kuhn–Tucker	74
KO	$k$ -optimality	37
L-PCA	linear PCA based objective reduction	94
LSSC	least squares stopping criterion	74
HypE	hypervolume estimation algorithm for multi-objective optimisation	13

IBEA	indicator-based evolutionary algorithm	12
MaOP	many-objective optimisation problem	16
MCDA	multiple criteria decision-analysis	14
MCDM	multiple criteria decision-making	14
MGBM	Marti–Garcia–Berlanga–Molina	73
MDS	multi-dimensional scaling	65
MOEA	multi-objective evolutionary algorithm	3
MOEA/D	multi-objective evolutionary algorithm based on decomposition	13
MOGLS	multi-objective genetic local search	40
MOP	multi-objective optimisation problem	1
MSOPS	multiple single objective Pareto sampling	13
MSOPS-II	multiple single objective Pareto sampling II	13
MVU	maximum variance unfolding	23
NL-MVU-PCA	nonlinear MVU-PCA based objective reduction	94
NSGA-II	non-dominated sorting genetic algorithm II	12
OCD	online convergence detection	73
OFCD	offline convergence detection	73
PBEA	preference-based evolutionary algorithm	44
PCA	principal component analysis	23
PCS	Pareto corner search	56
PESA	Pareto envelope-based selection algorithm	12
PI-EMO-VF	progressively interactive EMO using value functions	20
POS	Pareto-optimal set	3
POF	Pareto-optimal front	3
RCM	reduced correlation matrix	98
SDP	semidefinite programming	67
SIBEA	simple indicator based evolutionary algorithm	13

SMS-EMOA	$\mathcal{S}$ -metric selection MOEA	13
SPEA2	modified strength Pareto evolutionary algorithm 2	12
SR	sum ranking	36
SRF	sum ranking front	36
SWGR	sum of weighted global ratios	34
SWR	sum of weighted ratios	34
TC	threshold cut	63
UFS	unsupervised feature selection	55
WAR	weighted average ranking	34
WFG	walking fish group test problems	105
WMR	weighted maximum ranking	34
ZDT	Zitzler-Deb-Thiele test problems	72



# List of Figures

1.1	Modular mechanism of the evolutionary algorithms, depicted by four generic components. Taken from [1] and edited. . . . .	7
1.2	Structure of the thesis. . . . .	30
2.1	Structure of the literature review chapter. . . . .	35
2.2	Illustrating that the presence of redundant objectives can hinder the search efficiency of an MOEA. The $\mathcal{N}_{\mathcal{NS}}$ corresponds to a population size of 200 solutions and it is generated after 2000 generations (one run). . . . .	49
2.3	Parallel coordinator plot for a four-objective problem represented by four non-dominated or incomparable solutions. Example taken from [2] and edited. . .	52
2.4	Relation graphs corresponding to problem represented in Figure 2.3 for selected objective subsets. Taken from [2] and edited. . . . .	53
2.5	Illustration of the UFS procedure when applied to an arbitrary eight-objective problem. Note that, objectives $f_5$ and $f_6$ are removed in the end of the procedure. Example taken from [3] and edited. . . . .	56
2.6	Addressing submanifolds: A problem for standard PCA (Figures 2.6b and 2.6c have been taken from [4] and edited). . . . .	65
2.7	Maximum variance unfolding approach applied to the Swiss roll dataset showing different stages of the unfolding process (taken from [5]). . . . .	66
2.8	Structure of the research gaps section. . . . .	78

3.1	Steps of the methodology adopted by this thesis and its connection with proposed objectives and corresponding chapters. . . . .	88
4.1	A spectrum of possible redundancies in data for two sets of measurements. The best fit line is indicated by a dashed line. . . . .	99
4.2	Highlighting the scope and efficiency of the proposed framework. . . . .	105
4.3	Non-dominated set against principal components: Use of PCA for feature selection demonstrated on DTLZ5(2, 3). . . . .	111
4.4	DTLZ5(3, 5): Parallel coordinate plot for $\mathcal{N}_{\mathcal{P}}$ . . . . .	117
4.5	DTLZ5(3, 5): Parallel coordinate plots for $\mathcal{N}_{\mathcal{NS}}$ and $\mathcal{N}_{\epsilon}$ (one run). . . . .	127
4.6	DTLZ5(3, 5): POF-representation versus POF-approximation. . . . .	129
4.7	DTLZ5(3, 5): Illustrating that the $\mathcal{N}_{\mathcal{NS}}$ (one run) obtained for the reduced problem, $\mathcal{F}_{\mathcal{T}} = \{f_3, f_4, f_5\}$ , conforms with the true POF. . . . .	130
4.8	DTLZ5(3, 5): Parallel coordinate plots for $\mathcal{N}_{\epsilon}$ , $\mathcal{N}_{\mathcal{NS}}$ and $\mathcal{N}_{\mathcal{R}}$ (one run). . . .	131
4.9	Redundant problems: On POF-representation by $\mathcal{N}_{\mathcal{P}}$ , $\mathcal{N}_{\mathcal{R}}$ , $\mathcal{N}_{\epsilon}$ and $\mathcal{N}_{\mathcal{NS}}$ . . .	133
4.10	Non-redundant problems: On POF-representation by $\mathcal{N}_{\mathcal{P}}$ , $\mathcal{N}_{\mathcal{R}}$ , $\mathcal{N}_{\epsilon}$ and $\mathcal{N}_{\mathcal{NS}}$ . . .	140
4.11	Non-redundant problems: On POF-approximation of $\mathcal{N}_{\mathcal{P}}$ , $\mathcal{N}_{\epsilon}$ , $\mathcal{N}_{\mathcal{NS}}$ and $\mathcal{N}_{\mathcal{R}}$ using parallel coordinate plots for DTLZ1(15), DTLZ2(15) and DTLZ4(15) problems. . . . .	141
4.12	Effect of population size for $\mathcal{N}_{\epsilon}$ with $\theta = 0.997$ on the convergence level between the objectives depicted on parallel coordinate plots. The results correspond to DTLZ1(15) and a population size ranging from 200 to 1000 solutions. . . . .	150
4.13	Multi-speed gearbox design problem: The plots correspond to one run of NSGA-II. . . . .	151
4.14	Highlighting the nonlinear and linear characteristics of gearbox design problem. Nonlinearity in $f_1$ – $f_3$ subspace. One NSGA-II run. . . . .	152

4.15	Storm drainage system problem: Parallel coordinate plots (normalised), corresponding to one run of NSGA-II. . . . .	154
4.16	Highlighting the nonlinear and linear characteristics storm drainage system problem. Linearity in $f_1$ - $f_3$ subspace. One NSGA-II run. . . . .	154
4.17	Splines problem: $f_1 = x^2 + 1$ , $f_2 = -x^2 + x + 3$ , and $f_3 = -(f_1 + f_2^3)$ , where $x \in [-2, 2]$ . The analysis is based on $\mathcal{N}_{\mathcal{P}}$ . . . . .	156
5.1	On the difference between the MOEA and MCDM approaches. . . . .	166
5.2	Proposed approach on introducing <i>objectivity</i> , <i>repeatability</i> , <i>consistency</i> and <i>coherence</i> in DM's preferences, in the context of MCDM based MOEAs. . . .	166
5.3	A schematic for the proposed machine learning based framework to facilitate <i>objectivity</i> , <i>repeatability</i> , <i>consistency</i> and <i>coherence</i> in DM's preferences. . . .	168
5.4	Radar waveform problem: Visual representation of the $\delta$ -MOSS analysis by the Approach-A (NL-MVU-PCA based framework), averaged over 30 $\mathcal{N}_{\mathcal{MS}}$ .	181
5.5	Radar waveform problem: Visual representation of the $k$ -EMOSS analysis by the Approach-A (NL-MVU-PCA based framework), averaged over 30 $\mathcal{N}_{\mathcal{MS}}$ .	182
5.6	Car side-impact problem: Parallel coordinate plots, corresponding to one run of NSGA-II. . . . .	187
5.7	Car side-impact problem: Evolution of objective preference-weights captured by the Approach-A (NL-MVU-PCA based framework), averaged over 20 $\mathcal{N}_{\mathcal{NS}}$ . The weights are split in subfigures due to varying scales. . . . .	192
5.8	Car side-impact problem: Visual representation of the $\delta$ -MOSS analysis, by the Approach-A (NL-MVU-PCA based framework), averaged over 20 $\mathcal{N}_{\mathcal{NS}}$ . .	193
5.9	Car side-impact problem: Visual representation of the $k$ -EMOSS analysis, by the Approach-A (NL-MVU-PCA based framework), averaged over 20 $\mathcal{N}_{\mathcal{NS}}$ . .	194
5.10	Storm drainage system problem: Parallel coordinate plots, corresponding to one run of NSGA-II. . . . .	196

5.11	Storm drainage system problem: Evolution of objective preference-weights captured by the Approach-A (NL-MVU-PCA based framework), averaged over 20 $\mathcal{N}_{\mathcal{NS}}$ . The weights are split in subfigures due to varying scales. . . .	200
5.12	Storm drainage system problem: Visual representation of the $\delta$ -MOSS analysis, by the Approach-A (NL-MVU-PCA based framework), averaged over 20 $\mathcal{N}_{\mathcal{NS}}$ . . . . .	201
5.13	Storm drainage system problem: Visual representation of the $k$ -EMOSS analysis, by the Approach-A (NL-MVU-PCA based framework), averaged over 20 $\mathcal{N}_{\mathcal{NS}}$ . . . . .	201
5.14	Work roll cooling design problem: Parallel coordinate plots (normalised), corresponding to one run of NSGA-II. . . . .	206
5.15	Work roll cooling design problem: Evolution of objective preference-weights captured by the Approach-A (NL-MVU-PCA based framework), averaged over 20 $\mathcal{N}_{\mathcal{NS}}$ . The weights are split in subfigures due to varying scales. . . .	208
5.16	Work roll cooling design problem: Visual representation of the $\delta$ -MOSS analysis, by the Approach-A (NL-MVU-PCA based framework), averaged over 20 $\mathcal{N}_{\mathcal{NS}}$ . . . . .	209
5.17	Work roll cooling design problem: Visual representation of the $k$ -EMOSS analysis, by the Approach-A (NL-MVU-PCA based framework), averaged over 20 $\mathcal{N}_{\mathcal{NS}}$ . . . . .	210
5.18	Illustration: (a) Two linear or identically nonlinear objectives result in a <i>linear</i> non-dominated front, and (b) two objectives with different degree of nonlinearity result in a <i>nonlinear</i> non-dominated front, where use of $L_1$ or $L_2$ norm will be inadequate to account for nonlinearity, and $x \in [0, 1]$ . . . . .	212
5.19	Highlighting: (a) the limitations of $L_2$ norm when the data occupies a low-dimensional submanifold, and (b) the remedy through unfolding (MVU). . .	213

5.20	Illustration: Unless all the objectives are (a) linear, or (b) identically non-linear, the DRP approach assumption of comparable $\delta$ across all the objectives, over the entire non-dominated front will not hold. For all three cases, $x \in [0, 1]$ . $p_i$ and $q_i$ show the variation in $f_1$ and $f_2$ , respectively, over different regions of the front. . . . .	215
5.21	Highlighting the assumptions of the DRP approach, relating to the linearity of the front and equal distribution of solutions. LF and NLF stand for a linear and nonlinear front, respectively. EDS and N-EDS stand for equally distributed and not-equally distributed solutions, respectively. . . . .	215
5.22	Comparison between solution generated on the true POF (referred to as $\mathcal{N}_{\mathcal{P}}$ ) and solutions obtained by $\epsilon$ -MOEA (referred to as $\mathcal{N}_{\epsilon}$ ) (one run): Parallel coordinate plots for DTLZ5(2, 10) and DTLZ2 <sub>BZ</sub> (10), highlighting the <i>noise</i> in $\mathcal{N}_{\epsilon}$ as opposed to $\mathcal{N}_{\mathcal{P}}$ . . . . .	218
5.23	A comparative view on Approach-A vis-à-vis Approach-B for $\delta$ -MOSS and $k$ -EMOSS analysis applied to DTLZ5(2, 10), based on $\mathcal{N}_{\mathcal{P}}$ and $\mathcal{N}_{\epsilon}$ . . . . .	219
5.24	A comparative view on Approach-A vis-à-vis Approach-B for $\delta$ -MOSS and $k$ -EMOSS analysis applied to DTLZ2 <sub>BZ</sub> (10), based on $\mathcal{N}_{\mathcal{P}}$ and $\mathcal{N}_{\epsilon}$ . . . . .	220
6.1	Different combinations of $\mathcal{R}$ and $\mathcal{F}_p$ for the DTLZ5(2, 5). . . . .	237
6.2	Application of the <i>online</i> Framework 4 to DTLZ5(2, 5) (one run): Figure 6.2a depicts the random initial population vis-à-vis the true POF. Figures 6.2b and 6.2c presents the POF-approximation at $N_W = 2000$ (0th iteration) and $N_W = 4000$ (1st iteration), respectively. Figures 6.2d and 6.2e show the evolution of the quality indicators along the generations, namely, $g$ for convergence and $I_s$ for diversity. The dotted vertical line highlights $N_W = 2000$ which indicates the application of the online objective reduction. . . . .	238
6.3	The interplay of the random number $\mathcal{R}$ and probabilistic objective sets $\mathcal{F}_p$ , corresponding to erroneous preference information. . . . .	240

6.4	DTLZ5(2, 5): POF-approximation obtained by the <i>online</i> Framework 4 and <i>offline</i> Framework 1, while handling erroneous preference information. Here, $N_W = 2000$ and $N_g = 2000$ for <i>online</i> and <i>offline</i> frameworks, respectively, is used. . . . .	243
7.1	Effect of population dispersion in the number of cells and their corresponding region. . . . .	253
7.2	Unique cell identification number attributed to each cell by Equation 7.7, where $n_{bin} = 5$ leads to 25 cells for a two-objective problem. Solution A is attributed the cell identification number 21, while solutions B and C are assigned 18. . . . .	255
7.3	Representation of intersection and non-intersection regions obtained for populations $P$ and $Q$ accordingly to their probability distribution functions. . .	259
7.4	Non-redundant test problems: Influence of $n_{bin}$ on the proposed <i>dissimilarity</i> measure ( $\mathcal{D}(p, q)$ ) comprising of $\mathcal{D}(p, q)_I$ (for the intersection region) and $\mathcal{D}(p, q)_Y$ (for the non-intersection region). Each plot in this figure is averaged over 10 different solution sets obtained from NSGA-II runs, corresponding to a population size of 200 and 1000 generations. . . . .	268
7.5	Redundant test problems: Influence of $n_{bin}$ on the proposed <i>dissimilarity</i> measure ( $\mathcal{D}(p, q)$ ) comprising of $\mathcal{D}(p, q)_I$ (for the intersection region) and $\mathcal{D}(p, q)_Y$ (for the non-intersection region). Each plot in this figure is averaged over 10 different solution sets obtained from NSGA-II runs, corresponding to a population size of 200 and 1000 generations. . . . .	269
7.6	CTP1 to CTP4: Performance evaluation of Algorithm 6 through the NSGA-II populations at $N_{gt}$ corresponding to $n_p = 2, 3, 4$ ; and, $\mathcal{D}(p, q)$ along with its mean and standard deviation, and a zoomed snapshot limited to the first 100 generations. . . . .	271

7.7	CTP5 to CTP7: Performance evaluation of Algorithm 6 through the NSGA-II populations at $N_{gt}$ corresponding to $n_p = 2, 3, 4$ ; and, $\mathcal{D}(p, q)$ along with its mean and standard deviation, and a zoomed snapshot limited to the first 100 generations. . . . .	272
7.8	ZDT1 to ZDT3: Performance evaluation of Algorithm 6 through the NSGA-II populations at $N_{gt}$ corresponding to $n_p = 2, 3, 4$ ; and, $\mathcal{D}(p, q)$ along with its mean and standard deviation, and a zoomed snapshot limited to the first 100 generations. . . . .	273
7.9	ZDT4 to ZDT6: Performance evaluation of Algorithm 6 through the NSGA-II populations at $N_{gt}$ corresponding to $n_p = 2, 3, 4$ ; and, $\mathcal{D}(p, q)$ along with its mean and standard deviation, and a zoomed snapshot limited to the first 100 generations. . . . .	274
7.10	BNH, FON, KUR and OSY: Performance evaluation of Algorithm 6 through the NSGA-II populations at $N_{gt}$ corresponding to $n_p = 2, 3, 4$ ; and, $\mathcal{D}(p, q)$ along with its mean and standard deviation, and a zoomed snapshot limited to the first 100 generations. . . . .	277
7.11	POL, SCH1, SCH2 and SRN: Performance evaluation of Algorithm 6 through the NSGA-II populations at $N_{gt}$ corresponding to $n_p = 2, 3, 4$ ; and, $\mathcal{D}(p, q)$ along with its mean and standard deviation, and a zoomed snapshot limited to the first 100 generations. . . . .	278
7.12	TNK and VNT: Performance evaluation of Algorithm 6 through the NSGA-II populations at $N_{gt}$ corresponding to $n_p = 2, 3, 4$ ; and, $\mathcal{D}(p, q)$ along with its mean and standard deviation, and a zoomed snapshot limited to the first 100 generations. . . . .	279
7.13	DTLZ1( $M$ ) and DTLZ2( $M$ ) where $M = 5, 25$ : Performance evaluation of Algorithm 6 through the $\mathcal{D}(p, q)$ measure along with its mean and standard deviation, and a zoomed snapshot limited to first 100 generations. . . . .	282

7.14	DTLZ3( $M$ ) and DTLZ4( $M$ ) where $M = 5, 25$ : Performance evaluation of Algorithm 6 through the $\mathcal{D}(p, q)$ measure along with its mean and standard deviation, and a zoomed snapshot limited to first 100 generations. . . . .	283
7.15	DTLZ5(2, 5), DTLZ5(2, 20) and DTLZ5(2, 50): Performance evaluation of Algorithm 6 through the NSGA-II populations at $N_{gt}$ corresponding to $n_p = 2, 3, 4$ ; and $\mathcal{D}(p, q)$ measure along with its mean and standard deviation, and a zoomed snapshot limited to the first 100 generations. . . . .	284
7.16	DTLZ5(3, 20), DTLZ5(5, 20) and DTLZ5(7, 20): Performance evaluation of Algorithm 6 through the $\mathcal{D}(p, q)$ measure along with its mean and standard deviation, and a zoomed snapshot limited to first 100 generations. For the particular case of DTLZ5(3, 20), the NSGA-II populations at $N_{gt}$ corresponding to $n_p = 2, 3, 4$ are also shown. . . . .	285



# List of Tables

2.1	Selected multi-objective optimisation stopping criterion algorithms. The algorithms are organised by order of publication where the bottom elements are the most recent publications. . . . .	76
4.1	Classification of proposed enhancements, over [6, 7] . . . . .	103
4.2	Handling of problem redundancy, over [6, 7] . . . . .	104
4.3	$g$ function for DTLZ and DTLZ5( $I, M$ ) test problems . . . . .	108
4.4	Derivations for $\mathcal{D}_{\mathcal{T}}$ (dispersal of solutions on the true POF) . . . . .	109
4.5	Correlation matrix $R$ : Use of PCA for feature selection demonstrated on DTLZ5(2, 3). . . . .	111
4.6	DTLZ5(3, 5): The $R$ and $K$ matrix with their corresponding eigenvalues and eigenvectors, for $\mathcal{N}_{\mathcal{P}}$ (one run) . . . . .	113
4.7	DTLZ5(3, 5): Iteration1 of L-PCA with $\mathcal{N}_{\mathcal{P}}$ (one run) . . . . .	114
4.8	DTLZ5(3, 5): Iteration1 of NL-MVU-PCA with $\mathcal{N}_{\mathcal{P}}$ (one run) . . . . .	116
4.9	DTLZ5(3, 5): The $R$ and $K$ matrix with their corresponding eigenvalues and eigenvectors, for $\mathcal{N}_{\mathcal{NS}}$ (one run) . . . . .	118
4.10	DTLZ5(3, 5): Iteration1 of L-PCA with $\mathcal{N}_{\mathcal{NS}}$ (one run) . . . . .	119
4.11	DTLZ5(3, 5): Iteration1 of NL-MVU-PCA with $\mathcal{N}_{\mathcal{NS}}$ (one run) . . . . .	121
4.12	DTLZ5(3, 5): The $R$ and $K$ matrix with eigenvalues and eigenvectors of $K$ Iteration2 of NL-MVU-PCA (one run) . . . . .	121

4.13	DTLZ5(3, 5): Iteration2 of NL-MVU-PCA with $\mathcal{N}_{\mathcal{N}\mathcal{S}}$ (one run) . . . . .	122
4.14	DTLZ5(3, 5): The $R$ and $K$ matrix with their corresponding eigenvalues and eigenvectors, for $\mathcal{N}_\epsilon$ (one run) . . . . .	123
4.15	DTLZ5(3, 5): Iteration1 of L-PCA with $\mathcal{N}_\epsilon$ (one run) . . . . .	124
4.16	DTLZ5(3, 5): Iteration1 of NL-MVU-PCA with $\mathcal{N}_\epsilon$ (one run) . . . . .	125
4.17	DTLZ5(3, 5): POF-representation by $\mathcal{N}_{\mathcal{P}}$ , $\mathcal{N}_\epsilon$ and $\mathcal{N}_{\mathcal{N}\mathcal{S}}$ . . . . .	127
4.18	Redundant problems: Results based on $\mathcal{N}_{\mathcal{P}}$ with $\theta = 0.997$ . . . . .	131
4.19	Redundant problems: Results based on $\mathcal{N}_\epsilon$ , $\mathcal{N}_{\mathcal{N}\mathcal{S}}$ and $\mathcal{N}_{\mathcal{R}}$ with $\theta = 0.997$ . .	132
4.20	Redundant problems: on POF-approximation through convergence ( $g$ ) and diversity ( $I_s$ ) for $\mathcal{N}_\epsilon$ , $\mathcal{N}_{\mathcal{N}\mathcal{S}}$ and $\mathcal{N}_{\mathcal{R}}$ . . . . .	134
4.21	Errors ( $\mathcal{E}_T$ ) associated with NL-MVU-PCA and L-PCA when applied to $\mathcal{N}_\epsilon$ , $\mathcal{N}_{\mathcal{N}\mathcal{S}}$ and $\mathcal{N}_{\mathcal{R}}$ for redundant test problems, with $\theta = 0.997$ . . . . .	135
4.22	Non-redundant problems: Results based on $\mathcal{N}_\epsilon$ , $\mathcal{N}_{\mathcal{N}\mathcal{S}}$ and $\mathcal{N}_{\mathcal{R}}$ with $\theta = 0.997$	137
4.23	Non-redundant problems: on POF-approximation through convergence ( $g$ ) and diversity ( $I_s$ ) for $\mathcal{N}_\epsilon$ , $\mathcal{N}_{\mathcal{N}\mathcal{S}}$ and $\mathcal{N}_{\mathcal{R}}$ . . . . .	140
4.24	DTLZ1(15): Correlation-structure, eigenvalues and other highlights of L-PCA and NL-MVU-PCA corresponding to $\mathcal{N}_{\mathcal{P}}$ and $\mathcal{N}_\epsilon$ . . . . .	143
4.25	Errors ( $\mathcal{E}_T$ ) associated with NL-MVU-PCA and L-PCA when applied to $\mathcal{N}_\epsilon$ , $\mathcal{N}_{\mathcal{N}\mathcal{S}}$ and $\mathcal{N}_{\mathcal{R}}$ for non-redundant test problems, with $\theta = 0.997$ . . . . .	144
4.26	Customised framework for redundant problems: re-visiting DTLZ5(3, 5) cor- responding to $\mathcal{N}_{\mathcal{P}}$ . . . . .	146
4.27	Effect of the variance threshold ( $\theta = 0.954$ ) on the performance of the pro- posed L-PCA and NL-MVU-PCA algorithms on both redundant and non- redundant test problems, corresponding to $\mathcal{N}_{\mathcal{N}\mathcal{S}}$ and $\mathcal{N}_\epsilon$ . . . . .	148
4.28	Effect of the variance threshold ( $\theta = 0.682$ ) on the performance of the pro- posed L-PCA and NL-MVU-PCA algorithms on both redundant and non- redundant test problems, corresponding to $\mathcal{N}_{\mathcal{N}\mathcal{S}}$ and $\mathcal{N}_\epsilon$ . . . . .	148

4.29	Effect of population size for $\mathcal{N}_\epsilon$ with $\theta = 0.997$ on the performance of NL-MVU-PCA . . . . .	149
4.30	Performance of objective reduction algorithms on two real-world problems . . . . .	151
4.31	Splines problem: Correlation matrix ( $R$ ) for $\mathcal{N}_\mathcal{P}$ . . . . .	157
4.32	Splines problem: Key highlights of L-PCA and NL-MVU-PCA for $\mathcal{N}_\mathcal{P}$ . . . . .	157
4.33	Computational complexity of objective reduction algorithms . . . . .	159
5.1	Radar waveform problem: Approach-A (NL-MVU-PCA based framework), corresponding to $\mathcal{N}_{\mathcal{MS}}$ (one run) . . . . .	177
5.2	Radar waveform problem: $\delta$ -MOSS analysis by the Approach-A (NL-MVU-PCA based framework), averaged over 30 $\mathcal{N}_{\mathcal{MS}}$ . . . . .	180
5.3	Radar waveform problem: $k$ -EMOSS analysis by the Approach-A (NL-MVU-PCA based framework), averaged over 30 $\mathcal{N}_{\mathcal{MS}}$ . . . . .	180
5.4	Car side-impact problem: Approach-A (NL-MVU-PCA based framework), corresponding to $\mathcal{N}_{\mathcal{NS}}$ (one run) . . . . .	188
5.5	Car side-impact problem (Continuation from Table 5.4): Approach-A (NL-MVU-PCA based framework), corresponding to $\mathcal{N}_{\mathcal{NS}}$ (one run) . . . . .	189
5.6	Car side-impact problem: $\delta$ -MOSS analysis by the Approach-A (NL-MVU-PCA based framework), averaged over 20 $\mathcal{N}_{\mathcal{NS}}$ . . . . .	191
5.7	Car side-impact problem: $k$ -EMOSS analysis by the Approach-A (NL-MVU-PCA based framework), averaged over 20 $\mathcal{N}_{\mathcal{NS}}$ . . . . .	191
5.8	Storm drainage system problem: Approach-A based on $\mathcal{N}_{\mathcal{NS}}$ (one run) . . . . .	197
5.9	Storm drainage system problem: $\delta$ -MOSS analysis by the Approach-A (NL-MVU-PCA based framework), averaged over 20 $\mathcal{N}_{\mathcal{NS}}$ . . . . .	199
5.10	Storm drainage system problem: $k$ -EMOSS analysis by the Approach-A (NL-MVU-PCA based framework), averaged over 20 $\mathcal{N}_{\mathcal{NS}}$ . . . . .	199
5.11	Work roll cooling design problem: Approach-A based on $\mathcal{N}_{\mathcal{NS}}$ (one run) . . . . .	205

5.12	Work roll cooling design problem: $\delta$ -MOSS analysis by the Approach-A (NL-MVU-PCA based framework), averaged over 20 $\mathcal{N}_{\mathcal{NS}}$ . . . . .	208
5.13	Work roll cooling design problem: $k$ -EMOSS analysis by the Approach-A (NL-MVU-PCA based framework), averaged over 20 $\mathcal{N}_{\mathcal{NS}}$ . . . . .	208
5.14	$\delta$ -MOSS and $k$ -EMOSS analysis for the redundant DTLZ5( $I, M$ ) and non-redundant DTLZ2 <sub>BZ</sub> problems, based on $\mathcal{N}_{\mathcal{P}}$ . . . . .	220
5.15	$\delta$ -MOSS and $k$ -EMOSS analysis for the redundant DTLZ5( $I, M$ ) and non-redundant DTLZ2 <sub>BZ</sub> problems, averaged over 20 different samples of $\mathcal{N}_{\epsilon}$ . . . . .	221
6.1	Application of NL-MVU-PCA to DTLZ5(2, 5), corresponding to $N_W = 2000$ . . . . .	235
6.2	DTLZ5(2, 5): Performance of <i>online</i> Framework 4 versus <i>offline</i> Framework 1, while handling erroneous preference information . . . . .	240
6.3	Performance comparison between the <i>online</i> Framework 4 and <i>offline</i> Framework 1 with $\epsilon$ -MOEA as the underlying MOEA (data: $\mathcal{N}_{\epsilon}$ ) and NL-MVU-PCA as the underlying objective reduction algorithm . . . . .	245
6.4	On POF-approximation by $\mathcal{N}_{\epsilon-p-On}$ and $\mathcal{N}_{\epsilon-Off}$ , accessed by convergence ( $g$ ) and diversity ( $I_s$ ) measures . . . . .	245
7.1	Test-suite for the MOEA-termination algorithm (Algorithm 6). . . . .	266
7.2	$N_{gt}$ determined by the MOEA-termination algorithm (Algorithm 6) on two- and three-objective test problems, corresponding to $n_{bin} = 10$ , $n_s = 20$ and $n_p = 2, 3, 4$ . . . . .	275
7.3	$N_{gt}$ determined by the MOEA-termination algorithm (Algorithm 6) for many-objective versions of the DTLZ and DTLZ5( $I, M$ ) problems, corresponding to $n_{bin} = 10$ , $n_s = 20$ and $n_p = 2, 3, 4$ . . . . .	281
7.4	Highlighting the utility of the proposed MOEA-termination algorithm (Algorithm 6) for real-world problems. . . . .	288

# Chapter 1

## Introduction

An optimisation problem consists of minimising (or maximising) one or more objective function by systematically choosing input values, or parameters, from within an allowed set. In a single-objective optimisation problem only one objective function exists and therefore the best input values often correspond to only one single solution, also known as the optimal solution. However, when two or more objectives are considered together, the optimisation problem is known as a multi-objective optimisation problem (MOP), and conflicts might arise since an optimal solution that satisfies one of the objectives might not be optimal from the perspective of another objective. Therefore, when dealing with an MOP, instead of one optimal solution, there is a set of incomparable (or non-dominated) solutions often referred to as Pareto-optimal solutions. The Pareto-optimal solutions capture the conflicts existent between the objectives by representing the trade-offs that satisfy the objectives simultaneously. This means that a move from one Pareto-optimal solution to another implies a gain in one objective at the cost of a loss in another objective and when all objectives are considered together: (i) no Pareto-optimal solution is considered better than the other, and (ii) there are no other feasible solutions better than any Pareto-optimal solution [8].

## 1.1 Multi-objective Optimisation

A multi-objective optimisation problem deals with two or more objective functions which are meant to be maximised or minimised. When attempting to optimise the objective functions, a set of decision variables can take values within the decision (variable) space and the corresponding values mapped by the objective functions exist inside the so-called objective space. A *feasible* solution is then a set of decision variable values that satisfy a number of constraints, which can exist in both decision and objective space. Given this, a multi-objective optimisation problem (MOP) in its general form can be stated as:

$$\left. \begin{array}{ll} \text{minimise} & \mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_M(\mathbf{x}))^\top, \\ \text{subject to} & g_j(\mathbf{x}) \leq 0, \quad j = 1, \dots, J; \\ & h_l(\mathbf{x}) = 0, \quad l = 1, \dots, L; \\ & x_i^L \leq x_i \leq x_i^U, \quad i = 1, \dots, n. \end{array} \right\} \quad (1.1)$$

A solution is represented by the vector  $\mathbf{x}$  which is comprised of  $n$  decision variables as given by  $\mathbf{x} = (x_1, x_2, \dots, x_n)^\top$ . A decision variable  $x_i$  is bounded by the side constraints  $x_i^L$  and  $x_i^U$ , corresponding to lower and upper bound, respectively. The side constraints define the decision (variable) space  $\Omega$  and a solution  $\mathbf{x} \in \Omega$  is considered to be *feasible* only if it satisfies the side constraints and also  $J$  inequality and  $L$  equality constraints, given by  $g_j$  and  $h_l$ , respectively. Otherwise, the solution is called an *infeasible* solution. This means that not every solution in the decision space is feasible and the set of all feasible solutions is often known as simply the search space, and denoted here by  $\mathcal{D}$ . The objective space is defined by  $M$  objective functions and a solution  $\mathbf{x} \in \mathcal{D}$  can be represented by an  $M$ -dimensional objective vector, as given by  $f(\mathbf{x}) = \mathbf{z} = (z_1, z_2, \dots, z_M)^\top$ . This means between the variable space and objective space there is a mapping that translates an  $n$ -dimensional decision vector to an  $M$ -dimensional objective vector.

One important concept in multi-objective optimisation is known as dominance. In that,

let  $u$  and  $v$  represent two feasible solutions, i.e.,  $u, v \in \mathcal{D}$ . Then,  $u$  is said to dominate  $v$  if  $u_i \leq v_i$  for all  $i = 1, \dots, M$  and  $u \neq v$ . A solution  $\mathbf{x}^* \in \mathcal{D}$  is said to be a global Pareto-optimal solution if there is no  $\mathbf{x} \in \mathcal{D}$  that dominates  $\mathbf{x}^*$  such that  $\mathbf{x} \neq \mathbf{x}^*$ . The set of all Pareto-optimal solutions is known as the Pareto-optimal set (POS). The set of all the Pareto objective vectors, is often referred to as the Pareto-optimal front (POF) [9].

The POF dimensionality of an  $M$ -objective MOP is an  $m$ -dimensional manifold in the objective space where  $m$  is the number of conflicting (also critical or essential) objectives, such that  $m \leq M$ . The number of conflicting objectives is restricted by the number of free (unconstrained) decision variables, or by decision variables whose bounds are not too strict, that is, a sufficient degree of freedom is granted so that a curve is sampled along the objective space. Therefore, if there is only one free decision variable then only one curve can be sampled along the objective space and subsequently this curve can be approximated by a maximum of two conflicting objectives. This implies that for  $n$  free decision variables there could be a maximum of  $n + 1$  conflicting objectives, i.e.,  $m \leq n + 1$ . Hence, for  $n = 1$  the maximum number of conflicting objectives is two and the POF can be 2-dimensional, while for  $n = 2$  the maximum number of conflicting objectives is three and the POF can be 3-dimensional.

In multi-objective optimisation several approaches exist that can handle MOPs. The existing approaches can be categorised as traditional or evolutionary approaches. The traditional approaches in general use deterministic optimisation techniques to generate a single Pareto-optimal solution. On the other hand, the evolutionary approaches utilise stochastic optimisation techniques known as multi-objective evolutionary algorithms (MOEAs) that are able to generate a set of non-dominated solutions, aiming to approximate the complete POF. This means that solutions obtained by MOEAs are not necessarily Pareto-optimal solutions but rather an approximation. Since the MOEAs generate a set of solutions as opposed to traditional approaches that generate only one solution, it gives the opportunity to a decision maker (DM) to select a solution among the available options that pleases him/her

the most. In the subsections that follow the traditional and evolutionary approaches are described more in detail.

### 1.1.1 Traditional Approaches

Traditional approaches have been widely used to deal with MOPs and often they are not recognised as belonging to multi-objective optimisation. In general, these approaches convert the optimisation problem into a form where the objectives are all combined in a single-objective function. This allows the application of single-objective optimisation algorithms to solve MOPs. It was common to employ single-objective optimisation algorithms in dealing with MOPs before MOEAs were not yet available. However, nowadays it is preferable to employ MOEAs given their advantages over single-objective optimisation algorithms which will be mentioned further in this chapter.

An example of a traditional approach is the weighted sum [10, 11] as given by

$$\left. \begin{array}{ll} \text{minimise} & \sum_{i=1}^M w_i f_i(\mathbf{x}), \\ \text{subject to} & \mathbf{x} \in \mathcal{D}, \end{array} \right\} \quad (1.2)$$

where  $w_i \geq 0$  for all  $i = 1, \dots, M$ , and normally,  $\sum_{i=1}^M w_i = 1$ . To obtain more than one Pareto-optimal solution, several runs of a single-objective optimisation algorithm with different preference weight vectors are required. After the application of this approach several times, the DM is able to select the most desirable solution among the obtained Pareto-optimal solutions (*a posteriori* approach). Alternatively, as an *a priori* approach the DM can supply one or more preference-weight vectors before running the single-objective optimisation algorithm. In this way, each Pareto-optimal solution obtained will be of his/her preference. The popularity of this approach is driven by its usability and simplicity however the disadvantages that are associated with it cannot be ignored, such as:

- (i) as an *a priori* approach, the DM needs to represent his/her preferences in the form of



weight vectors. Hence, the DM requires an in-depth knowledge of the problem model in a global sense but also locally for different objective functions.

- (ii) as an *a posteriori* approach, an evenly distributed set of weights does not necessarily produce an evenly distributed representation in the decision (variable) space. As a consequence, some parts of the POF might be omitted and that might completely mislead the DM in regards to the available Pareto-optimal solutions.
- (iii) to obtain all Pareto-optimal solutions by alternating the weights it can only be achieved if the problem is convex. For non-convex problems the approach can easily get stuck in a sub-optimal solution.

Another commonly used traditional approach is known as  $\epsilon$ -constraint [12, 13] and is given by

$$\left. \begin{array}{ll} \text{minimise} & f_l(\mathbf{x}), \\ \text{subject to} & f_j(\mathbf{x}) \leq \epsilon_j \text{ for all } j = 1, \dots, M, j \neq l, \\ & \mathbf{x} \in \mathcal{D}, \end{array} \right\} \quad (1.3)$$

where  $l \in \{1, \dots, M\}$  and  $\epsilon_j$  are upper bounds for the objectives ( $j \neq l$ ). For this method, only one objective function is optimised while the remaining objectives are converted in to constraints. This approach is capable of dealing with convex and nonconvex problems which is considered to be an advantage over the weighed sum approach. Note that, the chosen upper bounds restrict the feasible region of the problem, therefore, if the upper bounds are not selected appropriately, one or more Pareto-optimal solutions might never be found since they exist in the infeasible region. Besides the two traditional approaches described above, other examples are: Benson's method [14], value function [15], and goal programming [16, 17, 18, 19].

### 1.1.2 Evolutionary Approaches

Evolutionary approaches are optimisation techniques that are increasingly popular for dealing with single- and multi-objective optimisation problems. In multi-objective optimisation these approaches are also known as multi-objective evolutionary algorithms (MOEAs) and the field of research is commonly referred to as evolutionary multi-objective optimisation (EMO).

The evolutionary term derives from Darwin's principle known as *natural selection* where the individuals of a population are selected based on their fitness. The selection takes place during an individual life cycle where reproduction is more accessible to the fittest. During reproduction (or recombination) the characteristics of the individuals are transferred to a new generation. As a result, the next generation is considered to be stronger and more adapted to the environment. When the concept is applied to the optimisation domain the individuals of a population are represented by solutions in the search space. Then, the solutions travel across the search space guided by the evolutionary operators towards the optimal solution(s).

The evolutionary based algorithms (including MOEAs) share a common modular mechanism, and in a generic way the components of the mechanism can be<sup>1</sup>: (i) population initialisation (ii) mating selection, (iii) variation and (iii) environmental selection. A representation of the components interactions is depicted in Figure 1.1 and each component is described as follows:

1. In the population initialisation component the solutions are sampled on the search space and in general a random function is utilised. This increases the chances for a population to cover different regions of the search space more effectively. The initial population is then referred to as the parent population.
2. The mating selection component aims to select a pre-defined number of promising

---

<sup>1</sup>The components have been suggested in [1] with the exception of population initialisation.

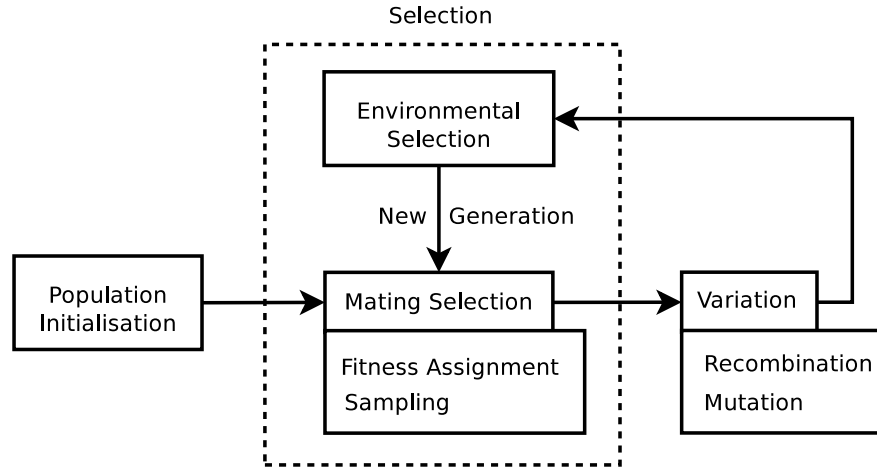


Figure 1.1: Modular mechanism of the evolutionary algorithms, depicted by four generic components. Taken from [1] and edited.

solutions from the parent population which are to be sent to a repository, known as the mating pool. To achieve this, the following two tasks are executed as follows:

- (a) the first task is known as *fitness assignment* and consists in the application of different metrics or heuristic functions to assign a relative merit of fitness to a particular solution. For instance, when a multi-objective problem is considered, the fitness function could combine all the objective functions in a single objective using weights to balance their importance. In this way each solution can be distinguished, based on their assigned fitness value.
  - (b) the second task is known as *sampling* and it consists of randomly picking solutions and selecting the best ones based on their fitness values. The best solutions are then sent to the *mating pool*. A commonly used sampling method is the binary tournament selection, where two solutions are randomly picked and the one with the best fitness value is selected.
3. The variation component applies two evolutionary operators to the solutions existing in the mating pool, namely recombination and mutation. In that:
- (a) the recombination operator combines parts of two solutions and new solution is

created. The operator is applied to a certain number of randomly picked solutions and a pre-defined number of solutions are created.

- (b) the mutation operator changes one or more characteristics of a specific solution, in an attempt to obtain some hidden advantage that might reveal to be beneficial in terms of fitness value.

The application of the recombination and mutation operators is generally governed by a user defined probability parameter, since constant recombinations and mutations might have a deteriorating effect in the population's overall fitness. After the application of the above two operators to the solutions in the mating pool, the new set of solutions is referred to as the child population.

4. The environmental selection component creates a new population using both the parent and child populations. Different strategies can be considered at this stage but the easiest one is to simply let the child population become the new population. However, it is recommended to preserve some good solutions that might exist in the parent population. Hence, the majority of the strategies take into account the parent and child populations while creating the new population, which is then referred to as the new parent population.

The above components are applied in an iterative manner as depicted in Figure 1.1, and at the end of each iteration a new population is created, leading to a new generation. This iterative process terminates when a specified termination criterion is satisfied. The criterion could be a predefined number of generations or until the solutions achieve a satisfactory fitness value.

### **Advantages over Traditional Approaches**

In many real-world problems the search space usually contains more than one optimal solution, due to nonlinearities and complex interactions among problem variables. While

attempting to solve these type of problems the traditional approaches when attracted to locally optimal solutions, often get stuck. Some examples of traditional approaches that work well for one type of problems but might not work well for others, are the geometric programming method or the Frank–Wolfe’s successive linear programming method [20]. The geometric programming method is only meant to solve problems with a *posynomial*-type objective function<sup>2</sup> and constraints and it cannot be easily applied to other type of problems. Alternatively, the Frank–Wolfe’s successive linear programming method is effective in dealing with a linear function and constraints. However, when applied to nonlinear problems the method is highly sensitive to the initial chosen parameters.

Given the above, the limitations of the traditional approaches when applied to MOPs, are as follows:

1. to effectively explore the search space one might need to conduct many single-objective runs, which can be very computational expensive, and a time demanding task;
2. in some methods if the problem is nonconvex not all Pareto-optimal solutions can be found;
3. all the traditional approaches require some problem knowledge, which is provided in the form of weights, objective bounds, or some other data structures;
4. the level of convergence obtained depends on the initial chosen parameters;
5. often a method that works well with one type of problem might not work well for other problems;
6. it is common for these methods to get stuck on local optimal solutions.

In particular, the traditional approaches without any adjustments might experience many difficulties when applied to a wide variety of optimisation problems. Hence, to counter the

---

<sup>2</sup>In a *posynomial* function the exponents can be arbitrary real numbers while its independent variables and coefficients must be positive real numbers. Not to be confused with *polynomial* functions where the exponents must be non-negative integers but its independent variables and coefficients can be arbitrary real numbers [21].

mentioned difficulties, the current trend is to employ evolutionary approaches as opposed to traditional ones. In that, some of the identified advantages [22, 8] of evolutionary approaches over traditional approaches are as follows:

- (i) conceptual simplicity: As depicted in Figure 1.1 any evolutionary strategy can be described in a general way using only four components. The solutions can be made to converge to the Pareto-optimal solutions only by iterating through the components.
- (ii) broad applicability: evolutionary approach can be applied to any problem as long it can be formulated as an optimisation task. The search space can be comprised of infeasible regions, and the possible solutions can be disjointed. The type of problems can be discrete combinatorial problems [23], mixed-integer problems [24, 25], continuous problems [26, 27, 28, 29], and many others.
- (iii) outperform traditional approaches on real-world problems: As previously mentioned, many real-world problems are characterised by many local optimal solutions, and the traditional approaches often get stuck on them. Since evolutionary approaches are population based methods and stochastic in nature, they are more robust and more capable of dealing with such problems.
- (iv) find multiple Pareto-optimal solutions in one single run: This is perhaps the most important advantage of evolutionary approaches over traditional ones. Lets say that after a single MOEA run, the outcome is a set of Pareto-optimal solutions. Each one of these solutions would correspond to the optimum (or a near-optimum) solution of a composite problem trading-off different objective functions. Therefore, each solution is important since it represents the existing trade-off relationships among the objectives. The problem involving multiple solutions is that a DM will have to decide which solution to pick. However, in the absence of such information it would be difficult for a DM to know what are the existing trade-offs among the objective functions. This is exactly the problem encountered in traditional approaches, where only a single solution is

generated at the end of a run. The solution so obtained is guided by the preference information provided by the DM. However, if the DM problem knowledge is poor or absent, the knowledge provided by multiple solutions (that represent existing trade-offs among the objective functions) is considered to be very important.

Despite their popularity, the existing evolutionary approaches are not foolproof and therefore some of their limitations/weaknesses (based on [30]) are as follows:

- (a) There is no guarantee that the global optimum will be found in feasible time. Although there are theoretical proofs of global convergence, these are only valid if infinite computation time is available and from a practical point of view that is never the case.
- (b) The parameters tuning is conducted mostly by trial-and-error. In that, one set of parameters might work well for a particular problem type but it is not guarantee that it will work for others. Therefore, a successful application of an evolutionary approach requires a lengthy trial-and-error procedure. To complicate more this matter, there is often a large number of parameters that need to be adjusted, such as: (i) the selection and crossover to use, (ii) population size, (iii) the probabilities associated with some operators and (iv) the form of fitness function.
- (c) The population approaches may be computational expensive. This is particularly the case for objective functions with high computational complexity. Hence, for such type of problems it is sometimes preferable to employ an application specific heuristic with domain-knowledge rather than an evolutionary approach.

### **State-of-the-art Multi-objective Evolutionary Algorithms (MOEAs)**

The aim of an MOEA is to approximate the POF of a given multi-objective problem where the objectives are often in conflict. During the optimisation task the MOEA attempts to:

- (i) improve convergence by minimising the distance of the solutions to the POF and, (ii) to improve diversity by maximising the spread of the solutions across the POF. To achieve

the above goals the strategies utilised by the MOEAs have to: (a) guide the search towards the POF, and (b) to keep the non-dominated solutions as diverse as possible. Towards this, several strategies have been proposed that can be classified as: (i) fitness assignment, (ii) diversity preservation or (iii) elitism. The fitness assignment strategies are used to improve convergence, while in contrast, diversity preservation improves diversity, as the name suggests. The elitism strategies are utilised to prevent good solutions from being replaced or deleted, hence, these strategies address both convergence and diversity. Based on the adopted strategies the existing MOEAs are now described.

The non-dominated sorting genetic algorithm II (NSGA-II) [31] is a well known MOEA that uses: (i) a Pareto-based fitness assignment strategy for convergence, and (ii) density information in the form of crowding distance for diversity. As an elitism strategy the new population is created based on parent and child populations. Many existing MOEAs present a framework similar to NSGA-II where two operators, one for convergence and the other for diversity, interchange iteratively. Other MOEAs that use a Pareto-based fitness assignment strategy are the SPEA2 [32], PESA [33], and  $\epsilon$ -MOEA [34]. As an elitism strategy these MOEAs use an external archive to preserve the best solutions and for diversity maintenance: (i) SPEA2 uses a nearest neighbour density estimation technique; (ii) PESA uses *hyperboxes* to divide the search space; and (iii)  $\epsilon$ -MOEA uses the  $\epsilon$ -dominance concept [35].

To measure the quality of the POF-approximation it is common to use indicators such as hypervolume [36] or generational distance [37]. The MOEAs that use the indicators information to guide the search process are known as indicator based MOEAs. The first to be suggested is the indicator-based evolutionary algorithm (IBEA) [38]. In IBEA, the fitness assignment strategy attributes to each solution a fitness value by calculating the “loss in quality” if a solution is to be removed. This is achieved by simply adding up the indicator values for each population member with respect to the rest of the population. The use of indicators to guide the search process improves the convergence and also the diversity, hence, no specific diversity maintenance strategies are necessary. Besides IBEA,



other existing indicator based MOEAs are:

- (i)  $\mathcal{S}$ -metric selection evolutionary multi-objective optimisation algorithm (SMS-EMOA) [39] where hypervolume governs the selection operator;
- (ii) simple indicator-based evolutionary algorithm (SIBEA) [40] that uses a modified hypervolume indicator with a weight distribution function; and
- (iii) hypervolume estimation algorithm for multi-objective optimisation (HypE) [41] that uses a fast method based on Monte Carlo simulations to estimate the hypervolume indicator.

The application of aggregation and target vector approaches is another strategy followed by some MOEAs. Examples are the MSOPS [42] and more recently the MSOPS-II [43]. As an elitist strategy, these MOEAs rank the solutions based on values obtained by the target vectors or aggregation function used, and a good spread of solutions is ensured by the target vectors' position. Another approach is to explore the traditional optimisation technique known as decomposition, to decompose a multi-objective optimisation problem in a number of scalar objective optimisation problems, called subproblems. This is the strategy followed by MOEA/D [44] where each subproblem consist of a (linear or nonlinear) weighted aggregation of individual objectives. The subproblems are optimised using the information from their neighbouring subproblems. The decomposition strategy is seen as addressing both convergence and diversity of the population.

Given that the MOEAs are stochastic methods, they can also be considered as sampling methods. In that sense, the solutions that they sample on the objective space, depending on the MOEA characteristics and also on the optimisation problem properties, can have: a uniform, Gaussian (more solutions close to the midrange than the extremes) or non-Gaussian (more solutions close to the extremes than midranges) distribution.

Moreover, although the above MOEAs also suffer from the general limitations/weakness of evolutionary algorithms (mentioned in the previous section), the more specific difficulties

that prevent them from successfully tackling MaOPs will be mentioned in Section 1.3.

## 1.2 Multiple Criteria Decision-Making (MCDM)

Multiple criteria decision-making (MCDM) or multiple criteria decision-analysis (MCDA) is an established research field that likewise EMO also deals with MOPs. However, in MCDM the aim is to help a *human* DM in finding a solution that best pleases him/her by considering the multiple criteria simultaneously. While in EMO, the aim is to generate a set of non-dominated solutions that best approximate the POF, and the intervention of the DM is not required during the process.

Traditionally, dealing with MOPs required the utilisation of mathematical optimisation techniques in alternation with decision-making, and the aim has been to give support to the DM in finding the solution that best fit his/her preferences. This continuous interchange between mathematical optimisation techniques and decision-making, created an interactive process that would lead to the most preferred solution by the DM. In this way the DM is deeply involved in the solution generation process which is guided by his/her preferences. During the decision-making process the DM's preferences are transformed into a type of preference model, which is then used to find the most preferred solution among the available alternatives, or to give indications that would guide the next optimisation task. Currently, many interactive methods exist and their differences are: (i) in the way that the DM is involved during the process, and (ii) in the format or model that is used to represent the preferences supplied by the DM [45]. The incorporation of the DM into methods that deal with MOPs has been the subject of many studies since the beginning of the 1970's [46, 47, 48]. However, many theoretical concepts did already exist, as in [49, 50, 51].

The DM is a person that is considered to be familiar with the given problem and he/she is also capable of providing preferences related to the objectives and/or to the solutions that are shown by the optimisation task. Another important aspect in the process is the

problem model itself. The problem model represents the knowledge of engineers and technical managers based on their joint expertise in the problem-domain. For a DM to elicit preference information and also to interpret better the information provided by the optimisation task, they are supposed to work closely with the engineers and technical managers that have modelled the problem.

Depending on the type of role of the DM in the solution generation process, the MCDM methods can be divided in four classes (as suggested in [9]): (i) no preference, (ii) *a priori*, (iii) *a posteriori* and (iv) interactive. A short description of each class is given below:

- (i) no preference: the solution is generated without any DM's preferences or any other type of preference information. Since there is no DM participation, the solution so obtained is neutral from the perspective of any preferences. Some available methods that belong to this class are the method of global criterion [52, 53] and the neutral compromise solution [54].
- (ii) *a priori*: the different aspirations and preferences supplied by the DM are first articulated before the optimisation process. The optimisation process is then guided by the provided information leading to a Pareto-optimal solution that satisfies the DM. Some available methods that belong to this class are the value function method [15], lexicographic ordering [55], and goal programming [16, 56].
- (iii) *a posteriori*: a set of Pareto-optimal solutions are initially generated and the DM selects the most desirable solution among the different alternatives. This allows the DM to get an overview of the problem's available solutions which could be very useful if the problem knowledge is not available or is very limited. Some available methods that belong to this class are the method of weight metrics [53] and the achievement scalarising function approach [57, 58].
- (iv) interactive: here decision-making and optimisation phases interchange, allowing the DM to participate more actively during the optimisation process. Initially the optimi-

sation phase presents to the DM some information (could be the best solution obtained so far). Then, the DM is asked to provide preference information in a format that can be used to guide the optimisation phase that follows. The process can be repeated until the obtained solution satisfies the DM's aspirations. During this process the DM is allowed to adjust the preference information as they learn the available alternatives that are created in each iteration. The interactive methods can be further divided into: (i) trade-off based methods, (ii) reference point approaches, and (iii) classification-based methods. A detailed review of methods that belong to each one of these categories is available in [59].

### 1.3 Many-objective Optimisation: Associated Difficulties

Optimisation problems with four or more objectives are often referred to as many-objective optimisation problems (MaOPs) [60, 61, 62]. There are several difficulties that influence the performance of MOEAs in tackling MaOPs as discussed in [63]. Fundamentally, the number of solutions required to give an approximation of the entire POF grows exponentially with a linear increase in the number of objectives. The dimensionality of the POF also suffers an increment together with the dimensionality of the objective space. In general, if the objective space dimensionality increases by one objective then the POF increases as well. In some situations the dimensionality of the POF can be inferior to the dimensionality of the objective space. If  $N$  points are needed in order to represent a one-dimensional POF, then  $O(N^M)$  points will be necessary to represent an  $M$ -dimensional POF, where  $M$  represents the number of objectives.

From the solution-approach perspective the traditional techniques require the DM to articulate preference-weights, so that the problem can be converted into single-objective by linear combination of objectives. However, when the number of objectives increases, the

articulation of preference-weights is though to be difficult due to:

1. research in psychology has shown that people are neither very good at handling precise and large amount of data or expressing them in quantified ways as desired for computational analysis [64, 65] nor they can handle more than several factors at once [66] and,
2. additionally, the visualisation of a POF and subsequently the selection of a solution with more than three dimensions requires the use of other techniques other than a simple Cartesian grid. Therefore, some sophisticated methods, such as the use of geodesic maps [67] or decision maps [68], have been considered, but they all require a huge number of points. For the DM, this means that the choice of a final solution is a challenging task.

From MOEAs perspective the known difficulties are:

1. domination-based MOEAs such as SPEA2 [32] and NSGA-II [31] rely on non-domination operator in order to rank the solutions. The number of non-dominated solutions in a population increases proportionally with the number of objectives, since there are many objectives where a trade-off (one is better in one or more objectives but worse in any other objective) can happen. At some point the number of non-dominated solutions is incrementally extended to the majority of the population. The efficacy of these algorithms to rank the population is therefore affected. Simultaneously, the selection pressure necessary to improve the population convergence towards the POF becomes weaker. In cases where the number of objectives is higher than five a random method is able to outperform NSGA-II as revealed experimentally in [62, 69]. Similarly, in [70], it is reported that the number of non-dominated fronts reduces (each being denser) with an increase in the number of objectives. Furthermore, the authors in [71] have argued that the concepts of Pareto-dominance and Pareto-optimality are not efficient in modelling or simulating human decision-making, as *optimality* is interpreted in a

narrow sense. The reasons presented include: (a) when comparing two solutions, a solution even marginally better in just one objective out of the many qualifies for non-domination; and (b) while comparing two solutions the human DM's process is not taken into consideration, such as: (i) how many objectives have improved, (ii) the size of such improvements, and (if there is any) (iii) the DM's preferences between objectives. Addressing these issues will lead to different concepts of *optimality* and several levels of dominance [72].

2. hypervolume calculation of a population where the objective number is higher than four is a computationally expensive task. In [73] a dimension-sweep algorithm for hypervolume computation achieves  $O(N^{M-2} \log N)$  time and linear space complexity in the worst-case, where  $M$  is the number of objectives and  $N$  the number of points (or solutions). Recently, the authors in [74] have achieved  $O(N^2)$  complexity but only in four dimensional problems. Hence, MOEAs that are based on hypervolume are inappropriate for dealing with MaOPs [75, 76]. Examples are IBEA [38], SMS-EMOA [39], SIBEA [40] and HypE [41]. Moreover, objective reduction techniques applied to SIBEA were considered in [77]. However, the authors have only demonstrated its effectiveness when applied to problems where all objectives are essential to describe the POF.
3. the decomposition based MOEA/D [43], requires the generation of weight vectors that are associated with each individual. The number of weights vectors is given by  $N = C_{M+1}^{H+M+1}$  where  $H$  is a user controlled parameter. If  $M \gg 4$  the number of weight vectors can become very high. To overcome this difficulty in [78] the authors have utilised different experimental design methods to generate more efficient weight vectors in high-dimensional space. However, the authors have not presented results for problems with more than five objectives.
4. subroutines and data-structures are also affected by the increase in the number of objectives; the growth in size and time complexity can be in some cases exponential [79].

For instance, in PESA [80] the  $M$ -objective fitness space is partitioned into *hyperboxes* to enable certain selection decisions. The above strategy works as follows. First the number of solutions occupying each hyperbox are estimated. Then the individuals in less explored regions (currently not occupied) are preferred. This in turn improves the population diversity. However, considering that each objective needs to be divided  $r$  times, the number of required hyperboxes is  $r^M$ . Hence, for problems with large  $M$ , the number of hyperboxes becomes impractical.

5. another reason for poor scalability in some algorithms is the term referred to as *fitness deterioration* in [81]. This issue, investigated in [79], occurs in algorithms with a size-restricted archive of points that evolves together with the population. When the archive is full, it is necessary to select points that will be replaced at some generation. However, some of the replaced points could dominate other points included in the archive, which would cause the population fitness to decrease. It is also expected to see this phenomenon getting worse in the many-objective cases. The authors have also pointed out that the adverse effect of fitness deterioration still remains to be investigated.

Having introduced the general difficulties that are associated with MaOPs, the following section highlights the importance of combining MCDM based approaches with MOEAs as a way to counter them. This relationship between MCDM based approaches and MOEAs is often referred to as MCDM based MOEAs.

## 1.4 MCDM based MOEAs for Many-objective Optimisation: Promise and Pitfalls

Given the above difficulties associated with MOEAs when dealing with MaOPs, there is an ever-growing interest in approaches that share elements of both EMO and MCDM based

approaches. These are commonly known as MCDM based MOEAs. First of all, in EMO as an *a posteriori* approach, an MOEA generates an approximate set of Pareto-optimal solutions and then the DM selects the most desirable solution. While in MCDM the DM's preferences can be incorporated before or during the solution search process corresponding to an *a priori* and an interactive approach, respectively. In the end of the solution generation process there is only one Pareto-optimal solution that corresponds to the DM aspirations.

To explore the benefits of integrating MCDM approaches directly into an MOEA, some researchers have adopted the MCDM *a priori* approach concept as an attempt to find a crowded set of Pareto-optimal solutions near the preferred region or preferred solution. Some attempts in this direction are the cone-domination based MOEA [82, 8], biased niching based MOEA [83], reference point based MOEAs [84, 85], the reference direction based MOEA [86], and the light beam approach based MOEA [87]. When a preferred region or a reference solution is specified, the mentioned MCDM based MOEAs are capable of finding a number of solutions that satisfy the DM aspirations. The drawback is that the DM interacts only at the beginning of the MOEA run.

A more effective approach is to integrate the preference information during the MOEA run. In this way, there is no need to wait until a preferred set of Pareto-optimal solutions (in case of *a priori*), or the complete POF (in case of *a posteriori*) is generated. Towards it, the DM is involved periodically during the MOEA run as the iterations are under way. At the end of each iteration the DM can inspect the currently evolved solutions and decide whether or not to change the preference information. This approach is less time consuming and more flexible than the previous approaches and also, it gives to the DM a greater sense of control over the solution generation process. An example is the PI-EMO-VF [88, 89], where the DM's preference information is used to construct a polynomial value function. At each iteration a subset of non-dominated solutions from the current population is shown to the DM. Then, the DM provides a rank to each solution and based on that information a polynomial value function is constructed. In this way the DM is able to control the solution



generation process better.

Considering the above discussion the benefits of integrating the MCDM approaches into MOEAs are:

1. the requirements of large population is negated since it is not intended to approximate the whole POF but only a crowded set of Pareto-optimal solutions near the preferred region designated by the DM;
2. the DM's preferences induce a preference order over the non-dominated solutions, which counters the poor selection pressure of most MOEAs; and
3. the difficulties associated with decision-making are avoided since the solutions obtained are based on the DM's preferences.

However, research in psychology and cognitive sciences [66, 64, 65] has shown that people in general are not good at handling a large amount of data in a precise way, nor can they handle several ( $7 \pm 2$ ) factors at a time. Also, the decision analysis conducted by a DM is often embroiled in subjectivity [90, 91]. Since it is based on personal experience, memory, thoughts, thinking paradigms and psychological states. These revelations warn that when dealing with MaOPs, the DM's preferences may be characterised by lack of: (i) *objectivity* (a rational basis); (ii) *repeatability* (identical preferences for identical options); (iii) *consistency* (alike preferences across multiple interaction stages); (iv) and *coherence* (alike preferences by multiple DMs).

The mentioned limitations of the DM, when dealing with MaOPs, calls for a decision support capable of articulating the DM's preferences with rationality. The section that follows touches upon this issue.

## 1.5 Objective Reduction based Decision Support for MCDM based MOEAs

As previously mentioned, the DM's preferences may be characterised by lack of: (i) objectivity, (ii) repeatability, (iii) consistency and (iv) coherence. This in turn affects the ability of MCDM based MOEAs in dealing with MaOPs. Some of the mentioned limitations could perhaps be countered if problem-specific information is made available to the DM prior to the decision-making task. In fact, problem-specific information is incorporated into the problem model by engineers and technical managers and it might not be available to the DM. In some other situations the information retained by the DM might be incomplete or erroneous. It is therefore important to inform the DM in regards to available problem-specific information, which in turn can prove to be beneficial to the decision-making task.

### Problem-specific information

When dealing with an MaOP, a DM might ask the following questions before articulating his/her preferences: (i) are all objectives necessary? (ii) what is the importance of each objective relative to the others? (iii) can the number of objective be reduced by incurring in some error? (iv) can the previous information be easily visualised in a timely manner? ...

To answer the above questions the objective reduction approaches are explored in this thesis. In that, these approaches attempt to reduce the number of objectives that are involved in the  $M$ -objective problem by assuming the existence of *redundant* objectives. A *redundant* objective can be omitted from the problem formulation and at the same time the POF will remain unaffected. On the other hand, if the omission of an objective affects the POF then the objective is termed as *essential*. The aim of the objective reduction approaches is to identify the smallest set of conflicting (or *essential*) objectives with size  $m$  ( $m \leq M$ ) by operating on the non-dominated solutions, often generated by an MOEA.

The benefits of applying an objective reduction approach to an  $M$ -objective problem are obvious if the reduced number of objectives is: (i)  $m \leq 3$ , the existing MOEAs are

capable of providing a good POF-approximation and (ii)  $4 \leq m < M$ , the benefits are in terms of computational cost, search efficiency and decision-making. However, considering that most MOEAs fail to provide a good POF-approximation when addressing MaOPs, it is important to select a criterion that could make accurate and reliable inferences about the dimensionality and composition of the true POF. In that, the criterion adopted by some promising objective reduction approaches is as follows:

- (i) preserve the dominance relations: dominance relation preservation (DRP) [92, 93, 2] is an approach that employs this criterion. An objective is considered to be redundant if its omission does not change the dominance relations of the given non-dominated solution set. Otherwise, the objective is termed as essential.
- (ii) preserve the correlation-structure: the approaches that employ this criterion utilise machine learning techniques such as principal component analysis (PCA) [94], Correntropy PCA [95], or maximum variance unfolding (MVU) [5], for interpreting the possible conflict between objectives existent in the given non-dominated set. The approaches are known as: (i) the PCA based PCA-NSGA-II [6], (ii) the correntropy PCA based C-PCA-NSGA-II [7] and (iii) the MVU based MVU-PCA-NSGA-II [7].

Moreover, a feature that could influence the accuracy of an objective reduction algorithm is the existence of *noise* in the given solution set. *Noise*, in this context, is defined as the difference between the dominance relations or the correlation-structure of a solution set and the one obtained on the true POF. To minimise the effect that noise might have in the accuracy of the objective reduction approach it is desirable to employ some *de-noising* operation to the given solution set. This *de-noising* operation could perhaps increase the accuracy and reliability of the information provided to the DM, which could be beneficial to the decision-making task.

## 1.6 Implementation of the Objective Reduction based Decision Support

The implementation of the objective reduction based decision support, depending at which stage an objective reduction algorithm is applied during the optimisation process, can be classified as *offline* or *online* reduction. These two implementations are described as follows:

1. *offline*: in here the objective reduction approach is applied “after” an MOEA search process. In general an MOEA is run for a pre-defined number of generations, and a set of non-dominated solutions is obtained at the end of the run. Subsequently, an objective reduction approach is applied to the final obtained non-dominated solution set. This type of implementation is useful for decision-making purposes since it provides problem specific information to a DM, and besides, it is very simple to implement.
2. *online*: in here the objective reduction approach is applied “during” an MOEA search process. Currently, different schemes exist where an objective reduction approach is integrated into an MOEA search process. Examples are: (i) DRP approach integrated into SIBEA [2] and (ii) UFS approach integrated into NSGA-II [96]. This type of implementation aims to improve the computational efficiency and search ability of an MOEA during the search for the Pareto-optimal solutions.

Another important aspect of an objective reduction implementation is the possibility of being implemented as an iterative procedure. An example is the offline implementation of the correlation–structure based objective reduction algorithms in [6, 7]. In that, the procedure starts with the original objective set and the objectives are eliminated in each iteration, until the number of objectives cannot be reduced further. If the number of objectives reduces between two iterations, the next MOEA run is able to provide a better POF–approximation than the previous iterations. Other examples of iterative procedures are the online implementations of the DRP and UFS approaches, proposed in [2] and [96], respectively.

The drawback of an iterative procedure is that an essential objective might be erroneously termed as redundant if the POF-representation of the solution set is poor, with no scope of being reconsidered. As a result, during the MOEA optimisation task it can happen that: (i) a part of the POF might be missing or (ii) if the remaining objectives are no longer in conflict, the population might collapse to a single Pareto-optimal solution. To prevent this situation, it is explored in this thesis the retention of all objectives so that an erroneously omitted essential objective can be recovered in subsequent iterations of the decision support framework.

## 1.7 *Timing* the Decision Support

During an MOEA optimisation process the population keeps changing along the iterations until it reaches the POF or until the search process stagnates. Independently of the chosen criterion for the decision support, it is expected for the correlation-structure or the dominance relations to change along the iterations until the population becomes stable, i.e, either converges to the POF or stagnates before convergence is achieved. In a case when the decision support is applied at an arbitrary MOEA generation it might happen that:

1. the analysis provided by the decision support might be inaccurate and unreliable if the population is not yet stable; and
2. several computational resources might have been wasted if more generations have been conducted beyond the stability instant.

To address the above two situations, it is important to determine when the population stabilises before applying the decision support. The benefits are:

- (i) the analysis provided by the decision support would be the most accurate and reliable as possible, given the MOEA capabilities; and

- (ii) the computational resources utilised by the optimisation process would be kept to a minimum.

To achieve this aim one needs to rely on a termination criterion capable of detecting when an MOEA population stabilises (either it converges or stagnates).

Termination criterion in multi-objective optimisation is a recent area of research. The available MOEA-termination algorithms utilise quality indicator as a way to measure differences between populations of solutions, which can be thought of as a kind of distance measure in multi-objective context. Some popular distance measures are: crowding distance [97], mutual dominance rate [98, 99],  $\epsilon$ -indicator [100], generational distance [101] and hypervolume [101, 100]. However, some difficulties have prevented the mentioned distance measures in dealing with MaOPs as follows: (i) the distance measures that rely on non-domination are not able to differentiate between solutions when the number of objectives increases; (ii) some distance measures computational complexity increases exponentially with an increase in the number of objectives; and (iii) distance measures that take the average of all individual distances make a stringent assumption that individual distances have to follow a Gaussian distribution.

The mentioned difficulties that are associated with available distance measures suggests that currently available MOEA-termination algorithms cannot effectively handle MaOPs. For an MOEA-termination algorithm to be able to handle MaOPs effectively it needs to:

1. detect changes between consecutive MOEA populations in high-dimensional space; and
2. yield a low computational complexity which allows the implementation of the criterion during an MOEA run.

Given the above, in this thesis, the application of information theory concepts such as *entropy* and *relative entropy* as an alternative to existing distance measures is explored. For this new measure to be successful in handling multi-objective optimisation problems it needs to be

able to detect during an MOEA run the following characteristics (as the case may be for a given problem, and a given MOEA):

- (a) to detect when the population converges to the POF, knowing that the majority of MOEAs are able to obtain a good POF–approximation when dealing with two- or three-objective problems.
- (b) to detect when the population stagnates, meaning that the MOEA is unable to further improve the quality of the POF–approximation despite additional computational expense. This type of situation is common when handling MaOPs, given that the majority of the MOEAs provide a poor POF–approximation

When the population of an MOEA stabilises, it is also expected for the criterion adopted by the decision support to stabilise as well. This is due to the fact that the solutions of a stabilised population (either converged to the POF or stagnated before convergence is achieved) are not expected to change if the MOEA is allowed to run for additional iterations. Hence, it can be inferred that the criterion adopted by the decision support which relies on those solutions, will remain unchanged. As a result, the benefits of the envisioned measure are not only meant for MOEAs but also for the decision support. In that, the accuracy and consistency of the information provided by the decision support could improve significantly.

## 1.8 Thesis Structure

The structure of this thesis is summarised below.

1. In Chapter 1 an introduction to the work developed in this thesis is provided and this includes: the traditional and evolutionary approaches in multi-objective optimisation, the fundamentals of multiple criteria decision-making (MCDM), the difficulties associated with existing MOEAs that prevent them from dealing with MaOPs efficiently, the promise and pitfalls of MCDM based MOEAs, objective reduction and its promise

as the basis for a decision support aimed at MCDM based MOEAs, concerns regarding implementation of the objective reduction based decision support, and why it is meaningful to apply the decision support at different optimisation instants.

2. A literature review is conducted in Chapter 2 that covers:
  - (a) the existing strategies meant to improve the low selection pressure for convergence that most MOEAs suffer from when dealing with MaOPs, with special focus on MCDM based MOEAs;
  - (b) the difficulties associated with human decision-making in handling more than  $(7 \pm 2)$  objectives at a time that affects the accuracy of existing MCDM based MOEAs when dealing with MaOPs;
  - (c) a review on existing objective reduction approaches focussing on available *offline* and *online* implementations;
  - (d) a review of existing termination criterion approaches that could indicate when it is more meaningful to apply the decision support.
  - (e) the research gaps that have been identified through the literature.
3. The aim and objectives which this thesis pursues are summarised in Chapter 3, along with the methodology adopted and the scope of the research.
4. A meaningful criterion for developing a decision support is identified in Chapter 4 and extended to be able to handle MaOPs with *noise* in the given solution set. Also, the selected criterion is compared vis-à-vis with an alternative approach on a wide range of test and real-world problems.
5. In Chapter 5 a machine learning based offline objective reduction framework is proposed. The proposed framework aims to introduce *objectivity*, *repeatability*, *consistency* and *coherence* in the DM's preferences to help him/her in dealing with MaOPs. The



benefits of the decision support can be shown for those MCDM based MOEAs that integrate the DM before or during the optimisation process.

6. In Chapter 6 an *online* implementation of the framework is proposed. The new implementation aims to address the limitations associated with the iterative procedure, found in the previously proposed *offline* objective reduction based decision support.
7. In Chapter 7 the fundamental question as to when the framework should be applied is addressed. Towards this, a termination criterion capable of detecting changes in high-dimensional space is proposed based on information theory concepts such as *entropy* and *relative entropy*.
8. The thesis concludes with Chapter 8, where the work developed is summarised based on: the contribution to knowledge, key findings, limitations and the scope for future work.

A diagram of the thesis structure is provided in Figure 1.2. The diagrams shows how the different chapters connect to each other. In that, Chapters 1 to 3 provide the introductory basis for the proposed decision support framework that aims to articulate the DM's preferences with rationality when dealing with MaOPs. The objectives pursued by the proposed framework are addressed in Chapters 4 to 7 separately. Then, in Chapter 8 the conclusions derived from the proposed decision support framework are discussed and summarised.

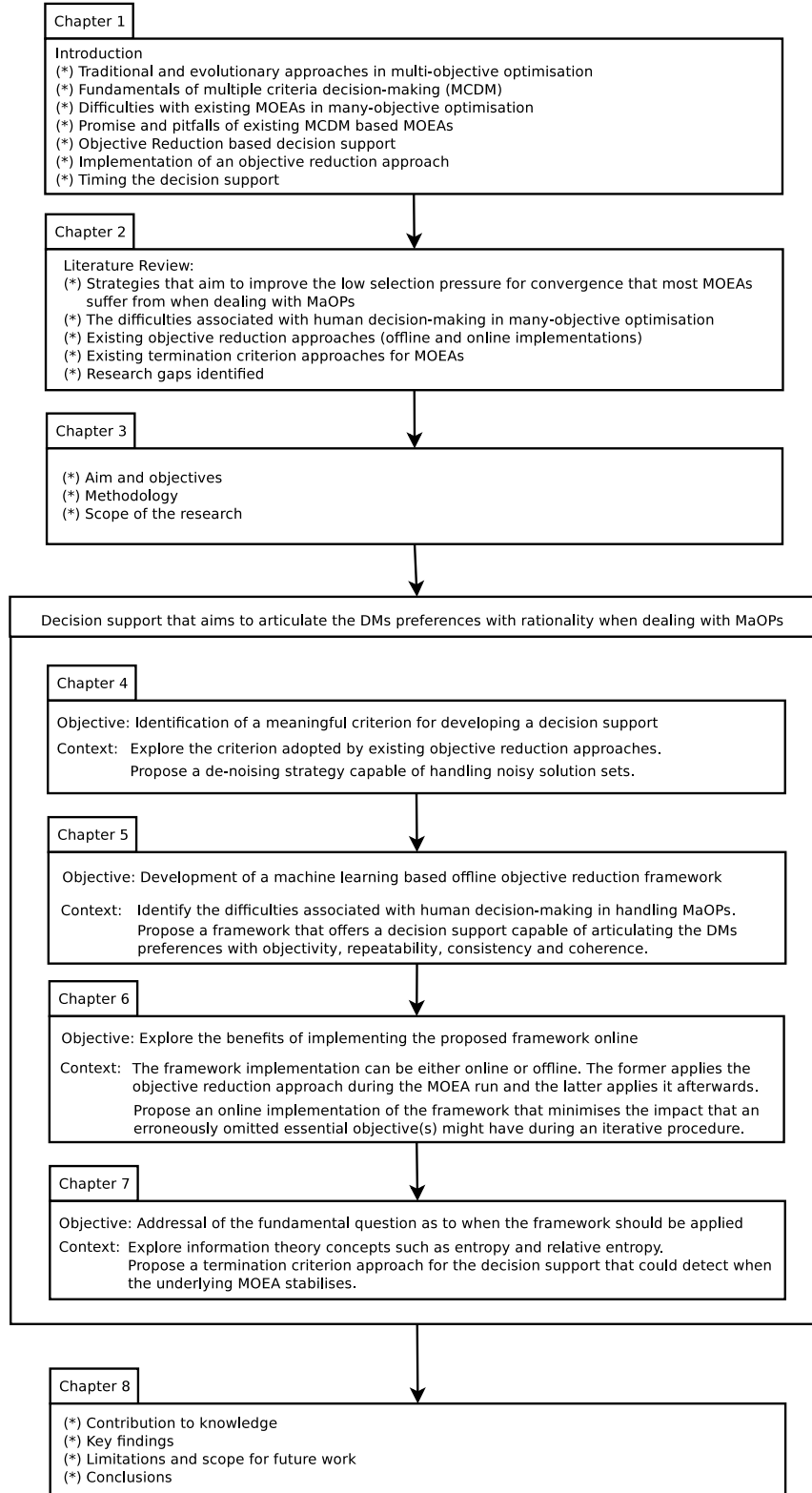


Figure 1.2: Structure of the thesis.

## Chapter 2

# Literature Review

The traditional approaches for solving multi-objective optimisation problems (MOPs) typically convert the multiple objectives into a single objective by scalarising the objectives. While this allows the application of widely available single-objective optimisation techniques, there is only one single Pareto-optimal solution at the end of the optimisation process. Therefore, the motivation for finding multiple Pareto-optimal solutions, in one single run, is one of the main reasons that made the multi-objective evolutionary algorithms (MOEAs) so popular in tackling MOPs [8]. However, when the number of objectives involved in an  $M$ -objective problem increases beyond three, that is  $M > 3$ , the majority of the MOEAs experience difficulties in approximating the Pareto-optimal solutions [63]. These optimisation problems where the number of objectives are higher than three are often referred to as many-objective optimisation problems (MaOPs) and can now be found in innumerable real-world applications [8, 102, 103, 104] and analytical test suites [105, 29, 27].

In an attempt to address the lack of efficiency that most MOEAs suffer from when dealing with MaOPs, the concept of Pareto-ranking is exploited in a more stringent way than Pareto-dominance by different strategies that induce a preference ordering over the non-dominated solutions [106]. Integrating these strategies into the MOEA's selection process will address some of the difficulties associated with Pareto-dominance. Depending on their role, they

attempt to: (i) assigning different ranks to non-dominated solutions, (ii) reduce the number of non-dominated solutions in the population, (iii) use scalarising fitness functions for fitness evaluation, (iv) use indicator functions to guide the search process, and (v) combine MCDM principles with an MOEA.

Although the above strategies are capable of improving the search efficiency of an MOEA, some accomplish it by improving convergence at the expense of diversity or the computational complexity increases exponentially with an increase in the number of objectives. Among the mentioned strategies, there is a growing interest in integrating the DM during the selection process by combining the MCDM based approaches with MOEAs [88, 89]. These approaches are here referred to as MCDM based MOEAs and their increasing popularity is mainly attributed to: (i) given that the aim is not to find the whole POF but rather a crowded set of solutions near the preferred area by the DM, the requirement of large population is negated; (ii) the low selection pressure for convergence that affects most MOEAs while differentiating between the non-dominated solutions is countered by the preferences of the DM; and (iii) the solutions obtained are based on the preferences of the DM therefore the difficulties associated with decision-making are avoided. Moreover, a more detailed description of the previously mentioned strategies and in particular the MCDM based MOEAs will be provided in Section 2.1 of this chapter.

The benefits of integrating the DM's preferences into an MOEA selection process are now clear, however, it is important to recognise that:

1. research in psychology and cognitive sciences [66, 64, 65] has shown that people in general are not good at handling a large amount of data in a precise way, nor can they handle several  $(7 \pm 2)$  factors at a time; and
2. the decision analysis conducted by a DM is often embroiled in subjectivity [90, 91], since it is based on personal experience, memory, thoughts, thinking paradigms and psychological states.

The above revelations warn that when dealing with MaOPs, the DM’s preferences may be characterised by lack of: (i) *objectivity* (a rational basis); (ii) *repeatability* (identical preferences for identical options); (iii) *consistency* (alike preferences across multiple interaction stages); (iv) and *coherence* (alike preferences by multiple DMs). A more detailed description of the limitations associated with *human* decision-making will be provided in Section 2.2 further in this chapter.

Given the above, a systematic method is needed to assist the DM to make rational decisions, referred to here as a decision support. Consequently, it would enable the MCDM based MOEAs to deal with MaOPs more effectively. Towards this, it is envisioned that the preference-structure of the different objective functions, inherent in the problem model itself, could be learnt and utilised towards a decision support for the DMs. To extract the preference-structure that is inherent in the problem model of a given  $M$ -objective problem, the application of objective reduction approaches [2, 3, 107, 7] is here explored. These approaches attempt to learn the preference-structure by operating on a set of non-dominated solutions often obtained by an MOEA. Given that the objective reduction approaches are promising candidates to form the basis of the envisioned decision support, Section 2.3 will focus specifically in objective reduction, and its organisation as follows:

- (i) to identify a meaningful criterion for the decision support, the existing objective reduction approaches are described in Section 2.3.1 and special focus is given to machine learning based objective reduction approaches in Section 2.3.2.
- (ii) an objective reduction approach can be either implemented *offline* or *online*. While the former applies the objective reduction approach “after” an MOEA run, the latter applies the approach “during” the MOEA run. The existing *online* strategies are described in Section 2.3.3 and some of their pitfalls are identified.
- (iii) given that the population keeps changing during an MOEA run and implicitly the chosen criterion as well, it is therefore important to identify when it is more meaningful

to apply the decision during an MOEA run. Towards this, in Section 2.3.4 the existing MOEA-termination algorithms are described and some of their difficulties in dealing with MaOPs are identified.

To conclude the literature review, the research gaps that have been identified are described in Section 2.4. Moreover, a pictorial representation of this chapter's structure is shown in Figure 2.1.

## 2.1 Focus on Evolutionary Many-objective Optimisation

### 2.1.1 Assigning Different Ranks to Non-dominated Solutions

Ranking a population is a strategy that helps MOEAs during the solution selection process, since it is important to identify the best individuals before recombination takes place. However, it is also one of the main difficulties associated with MOEAs when handling problems with large number of objectives. In that, when the number of objectives increases it becomes harder to differentiate between solutions. Therefore, different ranking strategies have been the focus of some studies.

For instance, the authors in [108] have considered different methods such as weighted maximum ranking (WMR), weighted average ranking (WAR), sum of weighted ratios (SWR), and sum of weighted global ratios (SWGR). The authors were interested in the use of these techniques in cases where *a priori* preferences existed for the objectives, which could be expressed as weights. Later in [79] the authors have extended the previous methods by elaborating on: (i) the application without the preferences (all weights were set to one), and (ii) considering only solutions in the Pareto front. In that, some of the methods considered are:

1. Average ranking (AR): is based on WAR and denoted by average ranking front (ARF)

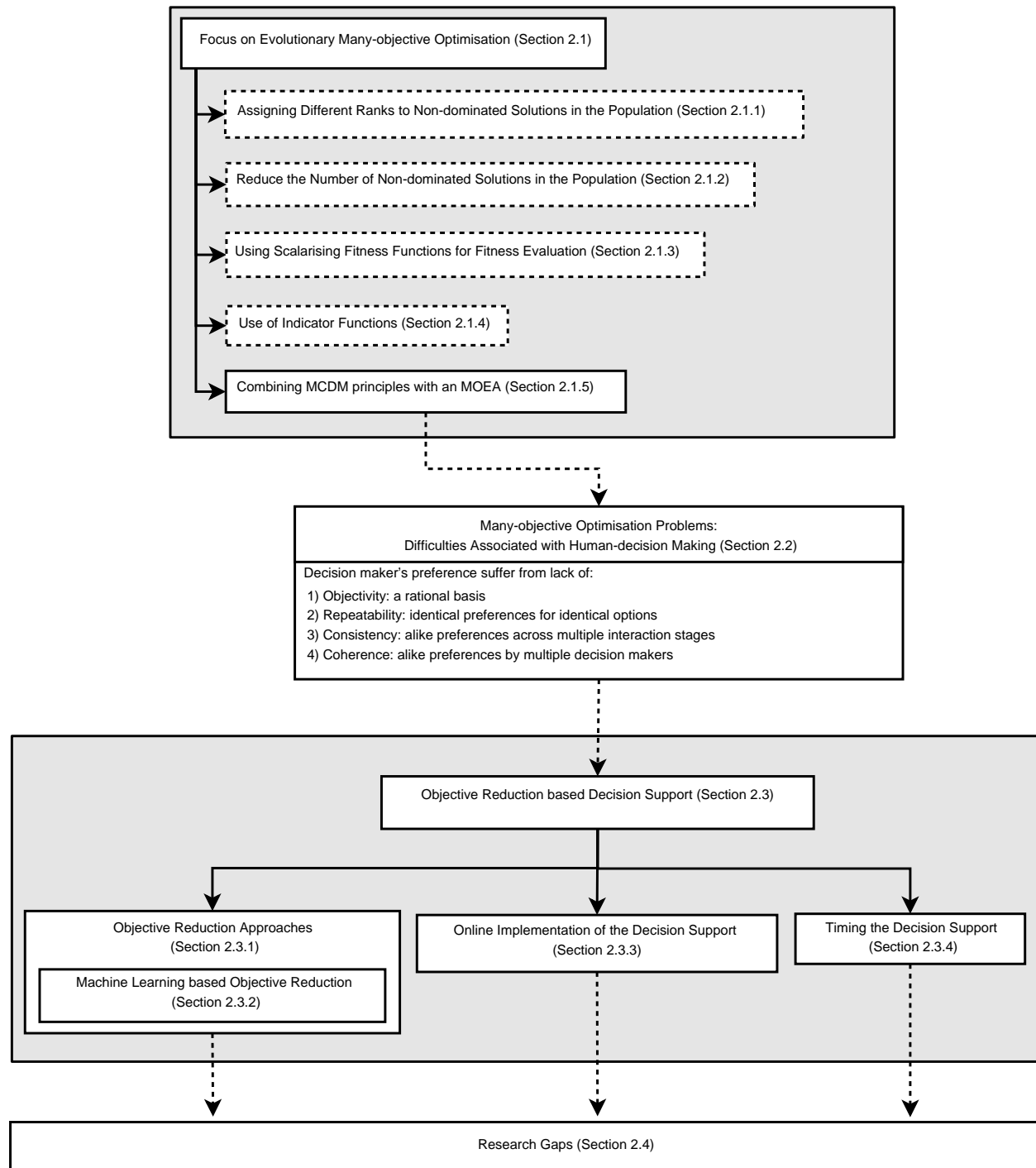


Figure 2.1: Structure of the literature review chapter.

in section “comparative experiments” [79]. The method attributes a score to each solution by summing the ranks of each objective. The ranks are obtained by ordering the solutions according to their objective values. For instance, let  $s$  and  $p$  be two non-dominated solutions for a three-objective problem. For  $s$ , if the 1st and 3rd objective are ranked 2 while 2nd is ranked 3 then  $AR(s) = 2 + 3 + 2 = 7$ . For  $p$ , if 1st and 2nd objective are ranked 2 while 3rd is ranked 5 then  $AR(p) = 2 + 2 + 5 = 9$ . Using the score attributed to  $s$  and  $p$  one can differentiate between them, and since  $AR(s) < AR(p)$  then  $s$  is preferable to  $p$ .

2. Sum ranking (SR): is based on SWR and denoted by sum ranking front (SRF) in section “comparative experiments” [79]. The only difference between this method and AR is the replacement of the objective rank by the normalised objective value. Mathematically, for a solution  $\mathbf{x}$  this is given by

$$SR(\mathbf{x}) = \sum_{k=1}^M \text{nratio}(\mathbf{x}, k),$$

where

$$\text{nratio}(\mathbf{x}, k) = \frac{f_k(\mathbf{x}) - \min_{\mathbf{x} \in P} \{f_k(\mathbf{x})\}}{\max_{\mathbf{x} \in P} \{f_k(\mathbf{x})\} - \min_{\mathbf{x} \in P} \{f_k(\mathbf{x})\}},$$

$M$  represents the number of objective and  $P$  is a set of non-dominated solutions.

3. Favour relation (FR) [109]: in this ranking scheme, a solution is considered better than another solution when the *favour* relation holds. The favour relation consists of the following: solution  $s$  is favoured over solution  $p$  only if  $s$  is better than  $p$  on more objectives, than on which  $p$  is better than  $s$ . This relation was later modified in [110] where the difference in objective values between two solutions have also been considered. One of the drawbacks of this approach is the lack of consideration by the objectives where  $p$  is better than  $s$ . This could, for instance, have some importance for the DM. Besides, it is also possible that the negative effect on these objectives is



strong enough to change the weighted sum metric completely.

4.  $k$ -optimality (KO) [111]: given that  $k$  represents the number of objectives, the authors have defined a relation between a non-dominated solution  $s$  and an efficiency order  $z$ , where  $(1 \leq z \leq k)$ . In this case, the relation says  $s$  is efficient of order  $z$  only *iff*  $s$  is non-dominated in every  $z$ -objective subset of the  $k$  objectives. When  $s$  is efficient of order  $z$  and at the same time  $z$  is the lowest possible value, then the authors consider  $s$  to be  $z$ -optimal. In order to obtain a rank between the non-dominated solutions, each solution  $s$  is associated with rank  $z$ , such that  $s$  is  $z$ -optimal.

A comparative analysis between the above methods, conducted in [79], revealed that ARF is the most successful ranking scheme as long as the objectives do not manifest a high inter-correlation value. In second place comes KO and FR in third. Additionally, it could be observed that FR works well only for problems with a low number of objectives.

Moreover, a different concept of *optimality* was introduced in [71]. This new concept consists of a fuzzy-based definition of optimality and dominated solutions. Different subsets of Pareto-optimal solutions could be obtained by considering: (i) a parameter that ranges from zero to one, and (ii) information provided by a DM. The method can be equal to the classical Pareto-optimality and dominance definitions by setting the parameter values as zero. Besides, the authors have also noted that the current definition of Pareto-optimality and Pareto-dominance are not adequate to model and simulate human decision-making. In that, the definitions do not account for: (i) how many objectives have improved, (ii) the size of such improvements and (iii) (if there is any) the DM's preferences between objectives.

### 2.1.2 Reduce the Number of Non-dominated Solutions in the Population

An approach called  $\epsilon$ -dominance [35] is used to control the population convergence by having a direct influence during the fitness assignment phase. This approach assumes the existence

of an  $\epsilon$  such that  $\epsilon > 0$ . Then, a solution  $\mathbf{x}$  is said to  $\epsilon$ -dominate solution  $\mathbf{y}$  (assuming maximisation) if

$$(1 + \epsilon) \cdot f_i(\mathbf{x}) \geq f_i(\mathbf{y}) \quad \forall_i \in \{1, \dots, M\}.$$

In practice the method divides the search space using a set of boxes with size  $\epsilon$ . Then only one non-dominated solution can be found inside each box. Depending on the chosen  $\epsilon$  different population ranks can emerge. In that, different  $\epsilon$ 's can have the following effect during an MOEA run:

- a higher  $\epsilon$  accelerates convergence but quality of the POF decreases; and
- in contrast a lower  $\epsilon$  produces a higher quality POF but with slower convergence.

Nonetheless, this approach eliminates boundary solutions, which in some problems may not be desirable, especially if the shape of the POF is unknown. Further investigation is also needed in finding an appropriate  $\epsilon$  value for different test problems since the different scales of the objectives need to be taken into account.

Another similar approach is the  $\alpha$ -dominance [112]. In this case, a solution  $\mathbf{x}$  is able to dominate solution  $\mathbf{y}$ , if, while being slightly inferior to  $\mathbf{y}$  in an objective, it is significantly superior to  $\mathbf{y}$  in the majority of objectives. Initially, for this approach a score vector needs to be computed, whose  $i$ -th component is given by

$$g_i(\mathbf{x}, \mathbf{y}) = f_i(\mathbf{x}) - f_i(\mathbf{y}) + \sum_{j \neq i}^M \alpha_{ij} (f_i(\mathbf{x}) - f_i(\mathbf{y}))$$

where  $f_i(\mathbf{x})$  represents the  $i$ -th objective value of solution  $\mathbf{x}$ . The variable  $\alpha_{ij}$  is a user defined parameter and sets the trade-off rate between objectives  $i$  and  $j$ . Then, a solution  $\mathbf{x}$   $\alpha$ -dominate a solution  $\mathbf{y}$  (assuming maximisation), if

$$g_i(\mathbf{x}, \mathbf{y}) \geq 0 \quad \forall_i \in \{1, \dots, M\}$$

and

$$\exists i \in \{1, \dots, M\} : g_i(\mathbf{x}, \mathbf{y}) > 0.$$

However, selecting an  $\alpha_{ij}$  appropriately is not an easy task. This requires an *a priori* problem knowledge that might not be available, such as: (i) the objectives' relative importance, (ii) how objective correlate and (iii) the problem POF structure.

Moreover, a method introduced in [113] is able to control the solution dominance area. This is achieved by creating an appropriate population ranking scheme. Unlike the above methods,  $\epsilon$ -dominance and  $\alpha$ -dominance that only support convergence, the new method is able to strengthen or weaken the selection process. Using a user-defined parameter  $S$ , the dominance area can be expanded or reduced. In some way this approach is different to conventional dominance methods, since modifying the dominance area promotes a different dominance relation by ranking solutions in a different way.

The authors while applying their scheme to 0/1 multi-objective knapsack problems have studied:

- the approach impact on search performance, by changing the complexity, size, and number of objectives handled by the multi-objective algorithm, and
- the effect of expanding the dominance area in the ranking of solutions.

As a result, the authors conclude that:

- convergence or diversity can be controlled by reducing or increasing the dominance area, and
- the optimal value in the area of dominance is directly affected by: (i) the complexity of the problem, (ii) size of the search space and (iii) the number of objectives in the problem.

### 2.1.3 Using Scalarising Fitness Functions for Fitness Evaluation

Scalarising fitness functions are often applied to single objective optimisation, where the objective function and fitness function are frequently identical. The same principle cannot be applied to multi-objective optimisation since multiple objectives have to be addressed simultaneously. Initially, to address this issue the fitness function takes the form of the weighted sum approach (Equation 1.2) which is given by

$$g^{ws}(\mathbf{x}, \mathbf{w}) = \sum_{i=1}^M w_i f_i(\mathbf{x}), \quad (2.1)$$

where the weight vector  $\mathbf{w} = (w_1, w_2, \dots, w_M)^\top$  often provided by a decision maker is normalised, i.e.,  $\sum_{i=1}^M w_i = 1$ . This method could be successfully integrated into an MOEA known as multi-objective genetic local search (MOGLS), first proposed in [114] and improved in [115]. The method reformulates the multi-objective problem by simultaneously optimising all weighted sum functions. Moreover, in [116] the weighted sum function could be integrated into NSGA-II using a probabilistic scheme. The obtained results show that the method outperforms original NSGA-II in terms of hypervolume measure.

A scalarising fitness function based MOEA is the Multiple Single Objective Pareto Sampling [42] (MSOPS) and more recent MSOPS-II [43], where the weighted min-max fitness function given by Equation 2.2 is used. As reported in [62], this approach is able to outperform NSGA-II when applied to MaOPs.

$$g^{wm}(\mathbf{x}, \mathbf{w}) = \max_{i=1}^M (w_i f_i(\mathbf{x})) \quad (2.2)$$

Another scalarising fitness function based MOEA is known as MOEA/D [44]. MOEA/D uses the traditional optimisation technique known as decomposition. The strategy decomposes the multi-objective optimisation problem into subproblems, where each subproblem has its own scalarising fitness function with different weight vectors. In this study besides

the weighted sum, two promising reference-based scalarising fitness functions have also been considered, namely: techbycheff [9] and normal-boundary intersection [117]. More recently the authors in [118] have proposed two schemes to integrate different types of scalarising function simultaneously into MOEA/D. In that, one scheme uses multiple grids of weights vectors and the other assigns different scalarising functions alternately to each weight vector in a single grid.

### 2.1.4 Use of Indicator Functions

Indicator functions have been used in multi-objective optimisation as a way to compare the performance of different MOEAs. These functions need to operate on a set of solutions since the outcome of an MOEA is not a single scalar value but instead a set of non-dominated solutions. Currently, there are many indicator functions (also known as quality measures) in the literature. For instance:

1. Some indicator functions measure the convergence of the solution set in approximating the POF. Some examples are: (i) the generational distance [37] that uses Euclidean distance to measure the distance between each solution and the closest solution known in the POF, and (ii) the error ratio [37] that counts the number of solutions that are Pareto-optimal.
2. Other indicator functions measure the diversity (or spread) of a solution set. Some examples are: (i) the maximum spread indicator [119] that measures the length of the diagonal of the hypercube with vertices set to the extreme objective values in the solution set, and (ii) the chi-square-like deviation measure [120] that uses probability theory concepts to measure the distribution of solutions in a sub-region of the non-dominated region.
3. There are some indicator functions that address both convergence and diversity. An example is the hypervolume [36] (also known as hyperarea metric [121] or S-metric [119]).

This indicator measures the volume delimited by a reference point and the solution set.

The use of indicator functions for fitness assignment was first explored by the indicator based evolutionary algorithm (IBEA) [38]. This MOEA uses an arbitrary indicator to compare pairs of solutions as a way to build a rank between them. Since the utilisation of indicator function addresses both convergence and diversity, hence, no additional convergence or diversity preservation approach is necessary. Another example is known as simple indicator based evolutionary algorithm (SIBEA) [40] where a weighted integration approach was used to extend the hypervolume indicator as a way to explore certain regions of the objective space. A more recent indicator based MOEA is known as HypE [41]. HypE uses a Monte Carlo simulation approach as an approximation to the exact hypervolume values. However, given that the computational complexity of indicators such as hypervolume increases exponentially with the number of objectives [122], their application to MaOPs is still considered impractical.

### 2.1.5 Combining MCDM principles with an MOEA

In the MCDM literature the multi-objective optimisation methods are classified into four classes accordingly to the DM role, as previously mentioned in Section 1.2. When there is no DM available it also means that there are no preferences as well. In this situation the methods are referred to as non-preference and the objective is to identify some neutral compromise solution. Given that the DM is available, the methods are classified as *a priori*, *a posteriori* or interactive. In an *a priori* method, the preferences are articulated by the DM before the solution generation process. Therefore, the successful application of an *a priori* method is dependent on the DM's problem knowledge. This could lead to biased solutions if the information retained by the DM is incorrect or absent. Alternatively, the solution generation process could provide the DM with a set of Pareto-optimal solutions and then utilise his/her preferences to select the most desirable alternative. These type of methods

are referred to as *a posteriori* and it is thought that the MOEAs belong to this category. There is a clear advantage over *a priori* methods since the set of Pareto-optimal solutions provides the DM with an idea of the available alternatives. However, for MOPs with a large number of objectives: (i) the DM might find it difficult to handle a large amount of information and (ii) the generation of a sufficient number of Pareto-optimal solutions might be too computationally expensive.

To address the drawbacks associated with *a priori* and *a posteriori* methods the DM availability can be explored in an interactive manner. In that, during the solution generation process the DM is provided with one or more solutions at the end of each iteration. This allows the DM to visualise the currently available alternatives during the optimisation process and at the same time to change his/her preferences at the end of each iteration. In this way, the DM feels more in charge of the overall process of generating the most desirable solution. The computational complexity of the interactive methods is considered to be low given that only the solutions desirable by the DM are generated.

The above discussion highlights the importance of integrating the DM's preferences during the solution generation process when attempting to solve MOPs. This not only avoids the time consuming operation of exploring the whole set of Pareto-optimal solutions but also provides to the DM a better knowledge of the problem being solved. However, the number of studies in the literature that have combined the DM's preferences with an MOEA as an interactive approach are still very low when compared with the MCDM literature.

The first attempt to introduce DM's preferences into an MOEA is reported in [123]. The authors have modified the fitness assignment method in a way that allows the DM to provide preference information in order to zoom into the POF region of his/her interest. Later, by using the concept of value functions to rank the population, the authors in [124] have integrated the DM's preferences during the environmental selection stage.

The authors in [125] have proposed a fuzzy approach to represent the DM's preferences in the format of *reference points*. The DM is asked to provide *reference points* until the

solutions of his/her interest are obtained. In [126] pairs of solutions are presented to the DM at the end of each iteration. Then, for each solution pair, the DM selects the solution of his/her interest based on their fitness values. The information provided by the DM is then utilised in the next iteration to generate new solutions that are based on his/her desirability.

The incorporation of preference information into NSGA-II lead the authors in [83] to modify the dominance and the crowding distance operator. Later, the idea of using *reference points* as a way to represent preference information could be successfully applied to NSGA-II in [84]. As a result, the strategy could find a user-controlled diverse set of solutions near the *reference points*. In [86] the concept of *reference direction*, first suggested in [127], could be successfully applied to NSGA-II. The strategy requests from the DM one of more *reference directions* that are used to guide solutions towards the region of interest. Furthermore, the concept of *light beam search*, first proposed in [128], was integrated into NSGA-II crowding distance operator in [87]. The concept consists in asking the DM by an aspiration and reservation point. These two points are then used to determine the direction of the search in an iteration.

A preference-based evolutionary algorithm (PBEA) based on IBEA was proposed in [85]. The strategy requests from the DM in each iteration for preference information in the format of a *reference point*. Then a new population is generated using that information by combining the fitness function and an *achievement (scalarising) function*<sup>1</sup>. In [129] the authors have explored the concept of *desirability functions* (DFs) which are used to replace the original objectives and are capable to express the preferences of the DM. The DFs have been integrated into the SMS-EMOEA.

Moreover, in [88, 89] the DM's preference information was used to construct a polynomial value function. The obtained value function was utilised to guide the search process during an NSGA-II run. In each iteration the strategy shows to the DM a subset of non-dominated solutions from the current population. The DM then ranks the solutions and the ranking

---

<sup>1</sup>An achievement (scalarising) function projects a *reference point* onto the set of non-dominated solutions. The components of the *reference point* represent the preference values of the objective functions.



provided is used to construct a value function considered to be representative of the DM aspirations.

### 2.1.6 Other Methods

A recent approach to induce selection pressure towards the POF has been proposed in [130] and is known as infeasibility driven evolutionary algorithm (IDEA). It is proposed the use of a controlled fraction of infeasible solutions (solutions close to the constraint boundary), as a way to induce the solutions from an MOEA population to convergence towards the POF. It is reported by the authors an improvement of around 98% in terms of convergence when compared with the original NSGA-II. For testing this concept the  $M$ -objective problem known as DTLZ5( $I, M$ ) [6], where  $I$  denotes the dimension of the POF, has been utilised. The only inconvenience of this approach is that it can only be applied to constrained problems where infeasible solutions can exist.

## 2.2 Many-objective Optimisation Problems: Difficulties Associated with *Human* Decision-making

The popularity of using MCDM based MOEAs is growing due to their ability to alleviate the difficulties associated with MaOPs, such as: (i) increasing number of solutions for higher  $M$ , (ii) poor search efficiency of most MOEAs, (iii) high-computational cost, and (iv) difficulties in visualisation. Thus:

- (a) the required number of solutions is reduced since the aim is not to find the whole POF, but rather, a set of solutions close to the preferred area desirable by the DM;
- (b) a preference order over non-dominated solutions induced by the DM's preferences improves the selection pressure and minimises the poor search efficiency of most MOEAs;

- (c) the difficulties in decision-making are reduced by the fact that the solutions are obtained based on the preferences of the DM.

The benefits of introducing the DM's preferences during an MOEA optimisation process are now clear. However, when the number of objectives involved in an optimisation problem is higher than three, as is case with MaOPs, the following needs to be noted in general decision-making:

1. Human capability of dealing with precise and large amount of data or expressing them in quantified ways are limited, as research in psychology has as shown [64, 65]. The number of factors that people are able to handle simultaneously are estimated to be  $(7 \pm 2)$  [66]. When dealing with many-objective problems it implies the following: (i) the information that is exchanged with the DM, taking into consideration the available multiple choices and the form that preferences are solicited, needs to be provided in the simplest way possible, and (ii) the DM will face difficulties in handling preferences for many-objective problems with 7 or more objectives.
2. The classical approaches from decision analysis and multiple criteria theory deal mostly with *subjective ranking* [91], as given by the following two arguments:
  - (a) decision making is based on “habitual domains” of the DM [90], that are constituted by: (i) personal experience, (ii) memory, (iii) thoughts, (iv) thinking paradigms and (v) psychological state. Due to the mentioned *cognitive limitations* of the DM, this argument might only hold true for multi-objective problems with two or three objectives but it will face difficulties for problems with higher number of objectives, such as 7 or more.
  - (b) people use the “perceived state of nature”, since there are precision limitations when performing measurements [91]. Hence, the “true state of nature” can, at best, be an approximation, which implies that full *objectivity* is not attainable.

In [131] the author has countered this argument and states that *objectivity* should not be dismissed but instead it can be seen as a constraint to achieving it.

3. In practice, when dealing with real-world problems, a single DM rarely exists. In [132] it is argued that in the decision-making process multiple DMs do exist and the whole process is redolent with feedback. When more DMs are involved there might be conflicting preferences and, as a result, finding the desirable alternatives becomes more difficult.
4. Decision-making is a dynamic process where pre-decision, decision and post-decision stages do occur and not just an act of selecting the most desirable alternatives [133].

The above arguments highlights the *subjectivity* and *cognitive limitations* of the DM, and also that during the decision-making process there could be more than one DM. In this context, it leads to lack of:

- (i) *objectivity*: preferences based on a rational basis, and not just based on “habitual domains” of the DM;
- (ii) *repeatability*: the same preferences for the same options;
- (iii) *consistency*: alike preferences across multiple interaction stages of the decision-making process;
- (iv) *coherence*: non-conflicting preferences by multiple DMs.

The above discussion highlights the necessity of a decision support for DMs to articulate their preferences with rationality when dealing with MaOPs. The decision support is promising for MCDM based MOEAs where the DM plays a crucial role during the optimisation process, leading to more accurate, consistent and rational solutions.

## 2.3 Objective Reduction based Decision Support

The objective reduction approaches aim to identify the smallest set of conflicting objectives which can generate the same POF as the one obtained by the *original* objective set, given by  $\mathcal{F}_0 = \{f_1, f_2, \dots, f_M\}$ . The smallest set of conflicting objectives can be also referred to as the *essential* objective set, given by  $(\mathcal{F}_\mathcal{T}, |\mathcal{F}_\mathcal{T}| = m)$ . The *dimensionality* of an  $M$ -objective problem can be described by the number of essential objectives, given by  $m$  ( $m \leq M$ ). Under the regularity condition [134], the dimensionality of a problem with  $m$  conflicting objectives is  $m - 1$ . However, in the current context, it is meant to refer to the number of essential objectives. After the essential objective set is identified for a given problem, the remaining set of objectives is referred to as the *redundant* objective set, given by  $\mathcal{F}_{redn} = \mathcal{F}_0 \setminus \mathcal{F}_\mathcal{T}$ . The redundant objective set can be omitted from the original problem and the POF should remain unaffected. Notably, a redundant objective can be considered as non-conflicting (or correlated) with some other objectives. While an essential objective is conflicting with one or more objectives.

The usefulness of objective reduction in recognising the redundant objectives is now shown using an example that involves two test problems and a non-dominated solution set generated by an MOEA. First, consider a non-dominated solution set generated by NSGA-II and denote it by  $\mathcal{N}_{\mathcal{NS}}$ . Second, consider the test problems with redundant objectives, namely DTLZ5(2, 3) and DTLZ5(3, 5), both instances of DTLZ5( $I, M$ ) [6]. The test problems POF is characterised as follows:

- (i) DTLZ5(2, 3): is a three-objective problem where objectives  $f_1$  and  $f_2$  are positively correlated and the essential objective can be either  $\{f_1, f_3\}$  or  $\{f_2, f_3\}$  with objective set size  $m = 2$ .
- (ii) DTLZ5(3, 5): is a five-objective problem where the objective set  $\{f_1, f_2, f_3\}$  are positively correlated and since, in that set, the variance of  $f_3$  is highest, the essential objective set can be described by  $\mathcal{F}_\mathcal{T} = \{f_3, f_4, f_5\}$  with  $m = 3$ .

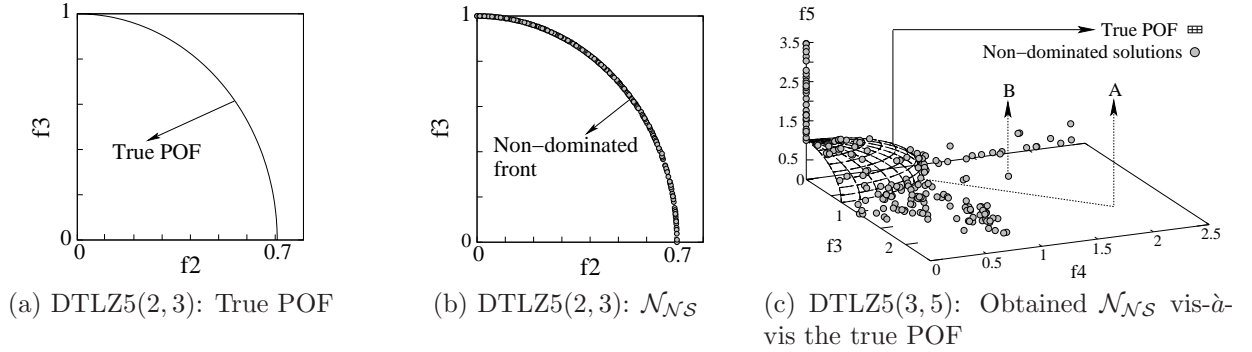


Figure 2.2: Illustrating that the presence of redundant objectives can hinder the search efficiency of an MOEA. The  $\mathcal{N}_{NS}$  corresponds to a population size of 200 solutions and it is generated after 2000 generations (one run).

It is known that NSGA-II is able to handle test problems with two or three objectives. However, it may struggle when the problem dimensionality is increased beyond four or more. The non-dominated set obtained by NSGA-II for DTLZ5(2, 3) as shown in Figure 2.2b is able to conform with the POF represented in Figure 2.2a. However, for DTLZ5(3, 5), where the number of essential objectives are only three, the non-dominated set could not converge to the true POF as evident in Figure 2.2c. It is also evident that the solution *A*, closer to the true POF, dominates solution *B* when only the objectives  $\{f_3, f_4, f_5\}$  are taken into account. Therefore, the reason for the existence of solution *B* is due to the presence of redundant objectives  $f_1$  or  $f_2$ . Solution *B* is considered non-dominated as long it is considered better in one of those objectives when compared with solution *A*. This example shows how redundant objectives can affect NSGA-II performance leading to poor convergence towards the POF. In a case when NSGA-II is allowed to run for infinite number of generations, the spurious points like *B* would be eventually dominated. However, for a timely reasonable number of generations those points would always prevail. While the previous example illustrates the importance of objective reduction, in general the benefits could be:

1. In a case when the number of objective in a problem is reduced to  $m \leq 3$ , an MOEA such as NSGA-II is able to converge to the true POF and solve a problem otherwise unsolvable.

2. In a case when an objective reduction is conducted but the number of essential objectives is  $4 \leq m \leq M$ , tasks such as decision-making, solution visualisation, and increase in MOEA performance due to higher computational efficiency can be improved.
3. The preference-based approaches could also benefit from dimensionality reduction since the problems that they are applied could be initially simplified.
4. When applied to real-world problems the knowledge about the problem can be improved by identifying possible correlated objectives in the formulation that otherwise may not be obvious.

### 2.3.1 Objective Reduction Approaches

Before describing the existing objective reduction approaches it is important to highlight some important salient issues that form the basis for existing approaches as follows:

1. The existing approaches operate on the non-dominated solutions generated by an MOEA.
2. If an objective reduction is performed after an MOEA run (*offline* reduction), the aim is to provide aid and support to the decision-making process. Or, if the objective reduction is performed during an MOEA run (*online* reduction), the aim is to simplify the search process.
3. An essential objective set can be expressed as either: (i) a subset of the problem original objectives or, (ii) as a linear combination of the original objectives. An objective reduction that pursues the former is a feature *selection* based approach, while if it pursues the latter then is a feature *extraction* based approach. Note that, the existing objective reduction approaches pursue feature *selection* since it is more straightforward to handle than feature *extraction* in terms of decision-making.

Given the above background, the existing objective reduction approaches are now presented in the following sections.

### Dominance Relation Preservation (DRP)

Dominance relation preservation (DRP) [92, 93, 2] is an objective reduction approach that takes into account the underlying dominance structure of the non-dominated solutions provided as input. Towards this, the authors utilise the notion of *weakly domination* where a solution  $\mathbf{x} \in X$  is said to *weakly dominate* a solution  $\mathbf{y} \in X$  iff  $\mathbf{x}$  is not worse than  $\mathbf{y}$  in all objectives. In respect to a particular objective subset  $\mathcal{F}'$  where  $\mathcal{F}' \subseteq \mathcal{F}$  and  $\mathcal{F}$  is the original objective set, this notion is given by

$$\mathbf{x} \preceq_{\mathcal{F}'} \mathbf{y} : \Longleftrightarrow \forall f_i \in \mathcal{F}' : f_i(\mathbf{x}) \leq f_i(\mathbf{y}), \quad (2.3)$$

and when generalised as a set notation for relations it is given by

$$\preceq_{\mathcal{F}'} = \{(\mathbf{x}, \mathbf{y}) \in A \times A \mid \forall f_i \in \mathcal{F}' : f_i(\mathbf{x}) \leq f_i(\mathbf{y})\}, \quad (2.4)$$

where  $A \subseteq X$  is a set of solutions. The notion of conflict between objectives is defined using the relation from Equation 2.4. Two objectives subsets  $\mathcal{F}_1$  and  $\mathcal{F}_2$  are considered in conflict if their weak dominance relation differs, that is,  $\preceq_{\mathcal{F}_1} \neq \preceq_{\mathcal{F}_2}$ , otherwise they are nonconflicting, meaning  $\preceq_{\mathcal{F}_1} = \preceq_{\mathcal{F}_2}$ .

To illustrate how the dominance structure can change when objectives are omitted, consider the example in Figure 2.3 that corresponds to a four-objective problem where four solutions are incomparable or non-dominated with respect to the original objective set, given by  $\mathcal{F} = \{f_1, f_2, f_3, f_4\}$ . The weak dominance relations can be represented by a *directed graph*, also called relation graph, where nodes represent solutions and an arrow between two solutions, say  $a$  and  $b$ , if  $a \rightarrow b$  then  $a$  weakly dominates  $b$ . The relation graphs for some

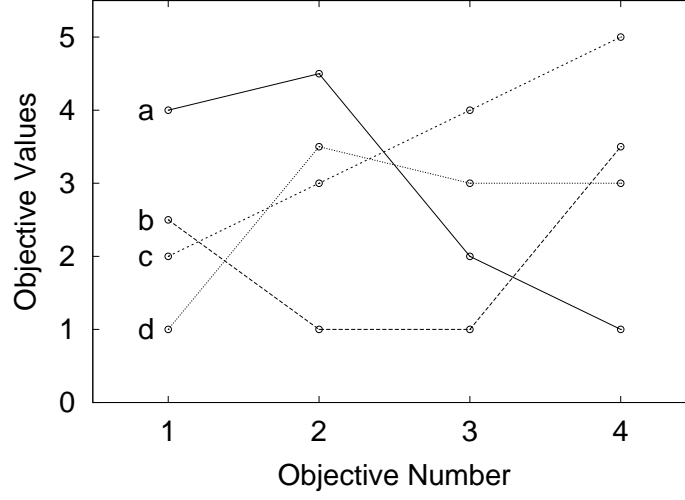


Figure 2.3: Parallel coordinator plot for a four-objective problem represented by four non-dominated or incomparable solutions. Example taken from [2] and edited.

objective subsets are depicted in Figure 2.4<sup>2</sup>. In that, consider the following observations:

- (a) In Figure 2.4a the dominance structure is captured w.r.t. the entire objective set. Notably, any solution weakly dominates itself, hence, there is always at least one arrow from a solution to itself. Since there is no solution that weakly dominates the others, the solutions are considered incomparable.
- (b) In Figure 2.4b the dominance structure is captured w.r.t.  $\{f_1, f_2, f_4\}$ . It is revealed that objective  $f_3$  is nonconflicting since its omission did not change the dominance structure when compared with Figure 2.4a, i.e.,  $\preceq_{\{f_1, f_2, f_3, f_4\}} = \preceq_{\{f_1, f_2, f_4\}}$ . This indicates that all solutions remain incomparable.
- (c) In Figure 2.4c the dominance structure is captured w.r.t.  $\{f_2, f_4\}$ . Notably, there is an arrow from solution  $b$  to solution  $c$  which means that  $b$  weakly dominates  $c$ , i.e.,  $b \preceq_{\{f_2, f_4\}} c$ . As a result, objective  $f_1$  is in conflict since its omission changed the dominance structure, i.e.,  $\preceq_{\{f_1, f_2, f_3, f_4\}} \neq \preceq_{\{f_2, f_4\}}$ . This also indicates that solution  $b$  is comparable and preferred to solution  $c$ .

<sup>2</sup>For a complete list of relation graphs corresponding to other objective subsets refer to [2].



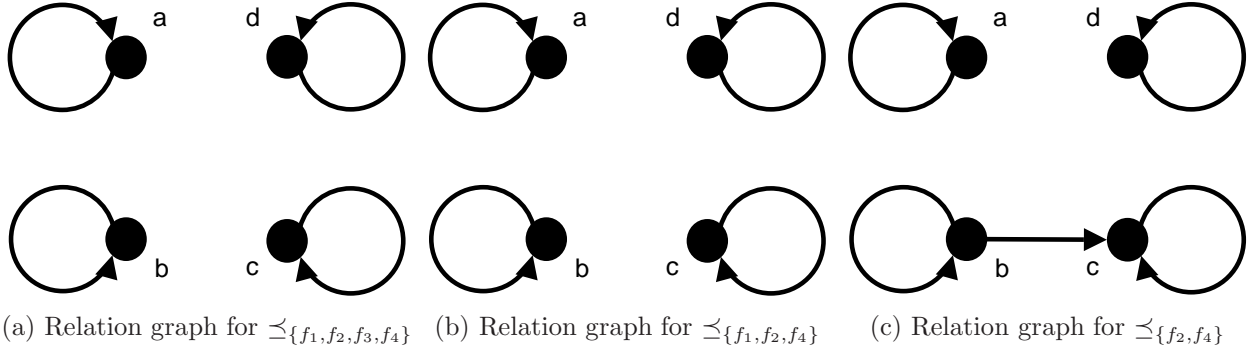


Figure 2.4: Relation graphs corresponding to problem represented in Figure 2.3 for selected objective subsets. Taken from [2] and edited.

From the above example note that:

- (i) objective  $f_1$  is a conflicting or *essential* since its omission changes the dominance structure of the original objective set, i.e.,  $\preceq_{f_2, f_4} \neq \preceq_{f_1, f_2, f_3, f_4}$ ;
- (ii) objective  $f_3$  is a nonconflicting or *redundant* since its omission does not change the dominance structure of the original objective set, i.e.,  $\preceq_{f_1, f_2, f_4} = \preceq_{f_1, f_2, f_3, f_4}$ ;

Moreover, by comparing the different relation graphs obtained for all possible objective subsets, one can determine the smallest objective subset that preserves the dominance structure of the original objective set. For the example in Figure 2.4, the objective set  $\{f_1, f_2, f_4\}$  is the smallest objective subset that preserves the dominance structure of the original objective set since its relation graph is equal to the one obtained by the entire objective set, i.e.,  $\preceq_{f_1, f_2, f_4} = \preceq_{f_1, f_2, f_3, f_4}$ .

Preserving the underlying dominance structure of the non-dominated solutions can be too stringent in practice, since the size of the minimal objective subset is often very close to the original objective set. To achieve a more substantial reduction of the objective set, the authors in [2] have proposed to quantify how much a solution should change its value, in the objective space, for the dominance structure between two objective subsets to remain unchanged. This concept can be used to reduce an objective set further in a case when a user is willing to incur in some error in order to deal with a smaller subset of objectives. In

this sense, the error that a user is willing to incur corresponds to the distance in objective space that a solution needs to be adjusted such that the dominance structure of the objective subset, i.e.  $\preceq_{\mathcal{F}'}$ , is equal to the one obtained by the original objective set, i.e.  $\preceq_{\mathcal{F}}$ .

To achieve this, it is necessary to consider all solution pairs and quantify the maximum error required to change their weak dominance relation until the dominance structure of two objective subsets matches. The maximum error required to change the dominance structure of an objective subset  $\mathcal{F}'$  ( $\mathcal{F}' \subseteq \mathcal{F}$ ) so that it matches with the original objective set  $\mathcal{F}$  is given by

$$\delta_{max}(A, \mathcal{F}', \mathcal{F}) = \max_{\substack{\mathbf{x}, \mathbf{y} \in A \\ \mathbf{x} \not\preceq_{\mathcal{F}} \mathbf{y} \\ \mathbf{x} \preceq_{\mathcal{F}'} \mathbf{y}}} \left\{ \max_{f_i \in \mathcal{F}} \{f_i(\mathbf{x}) - f_i(\mathbf{y})\} \right\}, \quad (2.5)$$

where  $A \subseteq X$  and  $\mathbf{x}$  and  $\mathbf{y}$  are two pairs of solutions. The rationale for the  $\delta_{max}$  is that when it is assumed that  $\mathbf{x}$  weakly dominates  $\mathbf{y}$  w.r.t.  $\mathcal{F}'$  it is also known that  $\mathbf{x}$  is not worse than  $\mathbf{y}$  for all objectives by an additive term  $\delta$ . As an example consider Figures 2.3 and 2.4c, where the error required to omit  $f_1$  and  $f_4$  by wrongly assuming that  $b$  weakly dominates  $c$  w.r.t.  $\{f_2, f_4\}$  and the entire objective set is  $\delta = 0.5$ , which is the distance between solutions  $b$  and  $c$  in  $f_1$ . Also, since the solution pair  $b$  and  $c$  is the only one depicted in Figure 2.4c that differs from the original objective set, the maximum error is then given by  $\delta_{max} = 0.5$  which is the quantified error that an user needs to incur to retain only  $f_2$  and  $f_4$  and discard  $f_1$  and  $f_3$  by preserving the dominance structure of the non-dominated solution set.

Using the previous definition of error one might be interested in determining the  $\delta$ -minimum objective subset, for short  $\delta$ -MOSS, which is the smallest objective set-retaining which would ensure that the error incurred is *just less than or equal to*  $\delta$ . Or one might want to determine the minimum objective subset of size  $k$  with minimum error, for short  $k$ -EMOSS. Towards this, using the above error definition the authors in [2] have proposed an exact and a greedy algorithm that are able to provide  $\delta$ -MOSS and  $k$ -EMOSS analysis.

### Unsupervised Feature Selection (UFS)

The Unsupervised feature selection (UFS) [3] is an objective reduction approach that uses the correlation matrix to measure the conflict between pairs of objectives. It is assumed that when two objectives increase or decrease together their correlation will be positive. On the other hand, the correlation will be positive if one objective decreases and the other increases, and vice versa. The more negative is the correlation between two objectives the more they conflict. The approach utilised by the authors is based on a technique for selecting a subset of original features from a given dataset prior to classification given in [135]. The conflict is measured using only the negative correlations between objectives, captured by  $1 - \rho(x, y) \in [0, 2]$ , where  $\rho(x, y)$  is the correlation coefficient between the features  $x$  and  $y$ . In this context the features  $x$  and  $y$  represent the objectives of the MOP. The objectives are considered completely positively correlated (nonconflict) if the result is zero and completely negatively correlated (conflict) if the result is 2.

The procedure followed by this objective reduction approach is as follows:

1. The objectives are divided in homogeneous neighbourhoods of size  $q$  that are formed around each objective. The distance between the objectives is determined by their correlation. Therefore, the higher the distance between two objectives the more they conflict. This step is demonstrated in Figure 2.5a for an eleven-objective problem where  $q = 2$  and two neighbourhoods are formed.
2. The most compact neighbourhood is selected, that is, the neighbourhood with the minimum farthest distance between two objectives. In Figure 2.5b the farthest neighbour is identified for each neighbourhood and it can be visually perceived that the neighbourhood in the left has the minimum farthest distance.
3. The centre of the neighbourhood is retained while the other  $q$  objectives are discarded. The error committed in removing the  $q$  objectives is equal to the distance of the farthest objective. In Figure 2.5c objectives  $f_5$  and  $f_6$  are removed and the error committed in

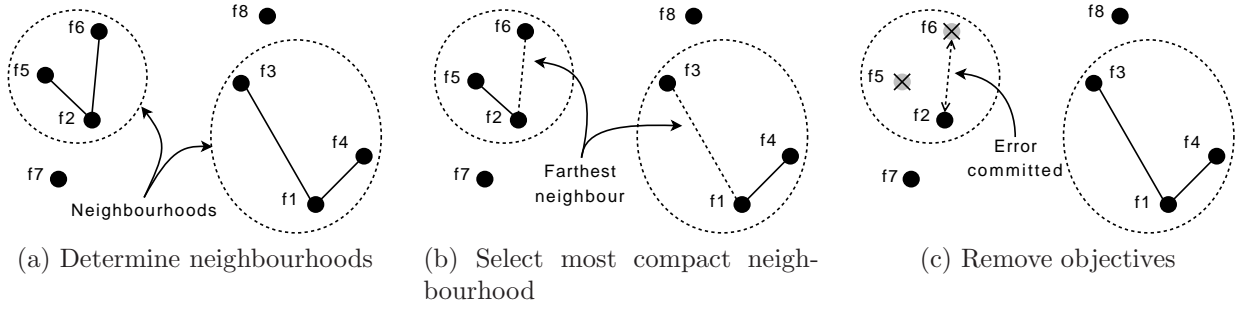


Figure 2.5: Illustration of the UFS procedure when applied to an arbitrary eight-objective problem. Note that, objectives  $f_5$  and  $f_6$  are removed in the end of the procedure. Example taken from [3] and edited.

this reduction is given by the distance between  $f_2$  and  $f_6$ , since their distance is the highest.

The procedure above described is repeated until a specified criterion is met and the neighbourhood of size  $q$  is reduced during the search. Moreover, using the above described procedure and their definition of error, the authors in [3] have proposed two algorithms that are capable of conducting  $\delta$ -MOSS and  $k$ -EMOSS analysis.

### Pareto Corner Search (PCS)

The Pareto corner search (PCS) [107] is an objective reduction approach that operates only with solutions in the corners of the POF, since it assumes that those solutions provide sufficient knowledge to capture the objectives' dependencies. Initially, the corner solutions that correspond to the POF are identified using a procedure similar to NSGA-II. This procedure attributes a higher rank to the solutions that are found in the corner of the Pareto-front, during the search process. In this way, these solutions have more chances of being selected over the generations and eventually it leads to the corner solutions of the POF. The objective reduction procedure that operates on the found (assumed to be) corner solutions of the POF, and identifies a redundant objective if its omission from the objective set has a insignificant effect over the number of non-dominated solutions on the population. Otherwise, the objective is identified as essential.

To measure the change in the number of non-dominated solutions the authors in [107] utilise a parameter  $R$  given by

$$R = N_{F_R - f_m} / N_F \quad (2.6)$$

where  $N_{F_R - f_m}$  is the number of non-dominated solutions that correspond to the objective set obtained after omitting  $f_m$  from the set  $F_R$  and,  $N_F$  is the number of non-dominated solutions that correspond to reference set  $F$ . The procedure starts by initialising the set  $F_R$  with the original objective set. Then, each objective is omitted from the set  $F_R$  and the change in the number of solutions is stored for each case. If the value measured by  $R$  is superior to a user defined threshold  $C$ , the objective is considered redundant. Otherwise, the objective is essential. This procedure is repeated until all objectives have been considered and the final  $F_R$  becomes the essential objective set of the problem in consideration. It is also highlighted that depending on the order that the objectives are removed, different essential objective sets might be obtained. To overcome this inconsistency the authors maintain at least one of the nonconflicting objectives in the essential set after considering the conflicts between various objectives.

### Machine Learning based Objective Reduction

This approach employs machine learning techniques such as principal components analysis (PCA) [94], correntropy PCA [95], and maximum variance unfolding (MVU) [5], for interpreting the possible conflict between objectives existent in the non-dominated solution set. The currently available algorithms are: (i) the PCA based PCA-NSGA-II [6], (ii) the correntropy PCA based C-PCA-NSGA-II and (iii) the MVU based MVU-PCA-NSGA-II. The last two algorithms have been proposed in [7]. Since this approach lies at the core of this thesis decision support framework, it is described further in Section 2.3.2.

### 2.3.2 Machine Learning based Objective Reduction

Machine learning based objective reduction employs techniques that assume the *intrinsic structure* of a *garbled* high-dimensional dataset can be revealed by transforming it such that the effect of *noise* and *redundancy* (dependencies) is minimised. PCA is able to achieve this by projecting a dataset  $X$  onto a lower dimensional linear space, known as the principal *subspace*, during this operation the correlation-structure of  $X$  is preserved [136]. More precisely, the task of PCA is to find a set of eigenvectors whose eigenvalues are the largest among those obtained from the correlation matrix of  $X$ , with the aim of decorrelating the data, i.e., to remove the second-order dependencies. PCA is therefore able to learn the existing second-order dependencies between solutions, and performs variance maximisation (or error minimisation) by assuming that the distribution in  $X$  is Gaussian. Given the above, PCA based objective reduction can be considered as a machine learning problem since PCA attempts to learn a subspace that captures the variations of the data. For clarity some of the terms mentioned above are described as follows:

- (i) The intrinsic structure of the POF is comprised by its intrinsic dimensionality ( $m$ ) and the essential components ( $\mathcal{F}_{\mathcal{T}}$ ).
- (ii) The *garbled* high-dimensional dataset refers to non-dominated solutions  $\mathcal{N}$  generated by an MOEA. The objective reduction techniques require the objective vectors of  $\mathcal{N}$  to operate on so that an essential objective set is determined. In that sense, in order to expect accurate results from an objective reduction technique the provided  $\mathcal{N}$  needs to be representative of the true POF. However, the  $\mathcal{N}$  obtained by most MOEAs is not representative of the true POF due to poor convergence and good diversity in areas of poor convergence. This explain why  $\mathcal{N}$  can be considered as *garbled*.
- (iii) *Unnoised signal* is represented by those non-dominated solutions that are considered to be on the POF. As an example, for DTLZ5(2,3), the  $\mathcal{N}_{NS}$  in Figure 2.2a totally

conforms with the POF. Therefore those non-dominated solutions can be considered *unnoised signal*. In the same way but for DTLZ5(3, 5) in Figure 2.2c solution *A* also conform with the POF, hence it can be considered *unnoised signal*.

- (iv) *Noised signal* is represented by those solutions that are not on the POF. For instance, for DTLZ5(3, 5) in Figure 2.2c solution *B* and all other solutions not on the POF are considered to be *noised signal*. *Noise* refers to the departure in the characteristics of the *noised* signal and those of the *unnoised* signal, for example, the difference in the dimensionality ( $m$ ) of *unnoised* signal and that of the *noised* signal (which could be greater than  $m$ ).
- (v) *Redundancy* represents the objectives that are non-conflicting (or correlated) with other objectives in the problem. It can also be inferred that *redundancy* may contribute to *garbled data*.

The existing machine learning based objective reduction algorithms are divided here in two categories, namely: (i) PCA Based Linear Objective Reduction and (ii) Kernel Based Nonlinear Objective Reduction. The former employs PCA to remove the second-order dependencies while the latter employs kernel based methods to remove the higher-order dependencies. In the following sections each category is described in detail.

### PCA Based Linear Objective Reduction

Principal component analysis (PCA), first presented in [137], is a simple and non-parametric technique widely used to extract relevant information from datasets hidden in an underlying structure by decorrelation, i.e., remove the second-order dependencies. It has been applied to diverse fields from neuroscience to computer graphics and depending on the field of application, it is also known as discrete Karhunen-Loève transform, the Hotelling transform or proper orthogonal decomposition.

When attempting to understand some phenomenon it is common to measure various

sources (e.g. temperature, velocity, humidity, etc.). During the data analysis when all the sources (or variables) are considered together, it might happen that the phenomenon looks unclear and some variables even redundant. To counter this situation, PCA uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The first principal component represents the highest variance possible that exists in the data and each succeeding component in turn represents the highest variance possible under the constraint that it needs to be orthogonal to the preceding components. The number of principal components is equal to the number of variables but depending on the amount of variance that a user wishes to retain, some principal components that represent lower variance can be discarded. In the section that follows a more mathematical description of PCA is provided.

### PCA — The Procedure

Consider a dataset, say  $X$  of dimensions  $M \times N$ , where  $M$  denotes the measurement types,  $N$  denotes the number of time samples and each time sample vector is designated by  $x_i \in \mathbb{R}^M$ . PCA is based on addressing the question as to how  $X$  could be best represented. It builds upon by considering a linear transformation, represented by a matrix, say  $V$  (with dimensions  $M \times M$ ), such that the new representation  $Y = VX$  will mean a change of basis for  $X$ .

PCA determines  $V$ , based on the properties desired in  $Y$ , which include: (i) high variance, and (ii) low redundancy. Both these properties can be achieved if the covariance matrix of  $Y$ , say  $C_Y$ , is diagonalised<sup>3</sup> implying that off-diagonal elements are zero. The process of diagonalising  $C_Y$  is restricted according to PCA assumption that the basis vectors from the linear transformation  $V$  must be orthogonal. It can be seen in [136] that diagonalisation of  $C_Y$  reveals that  $V$  is nothing but a matrix where each row (designated by  $v_j \in \mathbb{R}^M$ ) is an

---

<sup>3</sup>The diagonalisation process of a matrix, say  $A$ , consists in finding a diagonal matrix  $D = V^{-1}AV$  where  $V$  is an invertible matrix.



eigenvector of the covariance matrix (with dimensions  $M \times M$ ) of  $X$ , given by

$$C_X = \frac{1}{N} \dot{X} \dot{X}^T, \quad (2.7)$$

where  $\dot{X} = [\dot{x}_1 \ \dot{x}_2 \ \dots \ \dot{x}_N]$  and  $\dot{x}_i = x_i - \mu_{x_i}$ .

The eigenvectors are also called principal components, in that the eigenvector corresponding to the largest eigenvalue (largest variance) is referred to as the first principal component and the second largest eigenvalue (lower variance) is called the second principal component and so on. It is left up to the user to decide the amount of variance that he/she wishes to retain, in the transformed dataset  $Y$ , by selecting a reduced number of principal components that are kept in  $V$ . To conclude the procedure, the principal components ( $v_j$ ) are linearly combined with each time sample ( $x_i$ ) in order to form  $Y$ , given by

$$Y = \begin{bmatrix} v_1 \\ \vdots \\ v_M \end{bmatrix} \begin{bmatrix} x_1 & \dots & x_N \end{bmatrix} = \begin{bmatrix} v_1 \cdot x_1 & \dots & v_1 \cdot x_N \\ \vdots & \ddots & \vdots \\ v_M \cdot x_1 & \dots & v_M \cdot x_N \end{bmatrix}. \quad (2.8)$$

As an alternative to diagonalisation of the covariance matrix, PCA finds a linear transformation ( $V$ ) by minimising the mean square distance (error) between the dataset ( $X$ ) and the new representation ( $Y$ ) [137], as given by

$$\text{Minimise } Error(Y) = \frac{1}{N} \sum_{i=1}^N \|x_i - Vx_i\|^2, \quad (2.9)$$

or equivalently maximises the variance of the new representation ( $Y$ ) [138], as given by

$$\text{Maximise } Variance(Y) = \frac{1}{N} \sum_{i=1}^N \|Vx_i\|^2. \quad (2.10)$$

To summarise, PCA constructs a low-dimensional representation of the data by transforming it such that its global covariance structure is preserved, and the effect of *noise* and

*redundancy* is minimised.

### PCA — For Feature Selection

PCA is a classical feature extraction technique capable of reducing an original objective set  $\mathcal{F}_0 = \{f_1, f_2, \dots, f_M\}$  to  $\hat{\mathcal{F}} = \{\psi_1, \psi_2, \dots, \psi_m\}$  where  $m \ll M$  and each  $\psi_i$  for  $i = 1, \dots, m$  is a linear combination of  $f_j$ 's, where  $j = 1, \dots, M$ . The inherent challenge is to use PCA as a feature selection technique so that a subset of the given objectives describe the original problem better. In a traditional sense, PCA defines the new objectives as a combination of the original objectives, however such procedure pose difficulties in the decision-making process.

To overcome the mentioned limitation a PCA based procedure is proposed in [6], named PCA-NSGA-II, where the principal components are interpreted differently. The new interpretation selects the most important conflicting objectives from each principal component most-negative and most-positive elements. The rationale behind this procedure is summarised as follows:

1. The principal components can be understood as a directed ray in an  $M$ -dimensional space where each of the components can be seen as the direction cosines defining the ray.
2. The first component of a principal component vector, denotes the contribution of first objective towards this vector, the second element denotes the contribution of the second objective, and so on.
3. A positive value denotes an increase in objective value moving along this principal component (axes) and a negative denotes a decrease.

The procedure starts with analysing the first principal component and then proceeding to analyse the second principal component and so on, until all the significant components are considered. The principal components are analysed until the cumulative contribution

exceeds a pre-defined threshold cut ( $TC$ ). If a too high (close to 100%)  $TC$  is used, many redundant objectives may be chosen. On the other hand, if  $TC$  is too small then important objectives may be ignored. Since the objectives can have different scales the procedure instead of the covariance matrix ( $C_X$  defined in Equation 2.7) uses the correlation matrix as given by

$$R = \frac{1}{N} \ddot{X} \ddot{X}^T \quad (2.11)$$

where  $\ddot{X} = [\ddot{x}_1 \ \ddot{x}_2 \ \dots \ \ddot{x}_N]$  and  $\ddot{x}_i = (x_i - \mu_{x_i}) / \sigma_{x_i}$ . Using the correlation matrix standardises the data and prevents objectives with the highest variance to dominate the first principal component. To conclude, this procedure selects a subset of objectives from the original objective set by preserving the correlation-structure, maximising the variance and minimising the noise existent in the given non-dominated solution set.

### PCA — The Limitations

PCA-NSGA-II described above, could not successfully extract the correct problem dimensionality of all considered  $M$ -objective problems, as shown for test problems DTLZ5(2, 20), DTLZ5(5, 10) and DTLZ5(5, 20) in [6]. The reasons for this failure could be due to the limitations associated with PCA. As previously mentioned, PCA is applied to datasets when it is necessary to extract relevant information hidden in an underlying structure. In some cases PCA, “fails” to detect the correct answer and, for those cases, it is important to identify the responsible limitations. Given this, the assumptions on which the underlying PCA technique is based on, are considered as follows:

1. Assumption on linearity. PCA computes a different basis, in order to re-express the data, where the new data needs to be a linear combination of the basis vectors. This linearity assumption simplifies the problem by reducing the set of potential bases, its potential is clear when applied to a linear regime (Figure 2.6a). However its application to nonlinear regimes (Figure 2.6b) is not always possible.

2. Assumption that mean and covariance are sufficient statistics, i.e., removing second-order dependencies is sufficient at revealing all structure in the data. PCA assumes that  $Y$  can be characterised by the diagonalisation of  $C_Y$ , which is done considering only its mean and variance. However, only the Gaussian distribution can be fully described by its variance and clearly this assumption only holds for cases such as Figure 2.6a where the best fit line provided by the first eigenvector can fully describe the data structure. However, for non-Gaussian or multi-modal Gaussian data (Figure 2.6b), this assumption is understood to fail.
3. Assumption that principal components are orthogonal. Linear algebra decomposition techniques (e.g. eigenvector decomposition), are able to be applied to PCA since the principal components are orthogonal. This requirement specifies that all data must be arranged along orthogonal axes and, in some datasets, that is not the case, as shown in Figure 2.6c. A simple example is a plan of a big city where streets are normally connected to each other. The orthogonal assumption from PCA requires all the streets to be perpendicular or horizontal, which in some cases is far from reality.
4. Assumption that large variances have important dynamics. The strict procedure employed by PCA to select the structures with important dynamics is, in some cases, misleading. PCA assumes that there is redundancy in the data, implying that the principal components with large variance represent interesting structure, while those with lower variance represent noise. This assumption does not always work, especially if the data is non-Gaussian distributed. For instance, consider a case where a large portion of the variance is divided between non-orthogonal axes, then the axis with the largest variance does not correspond to the true interesting dynamics.

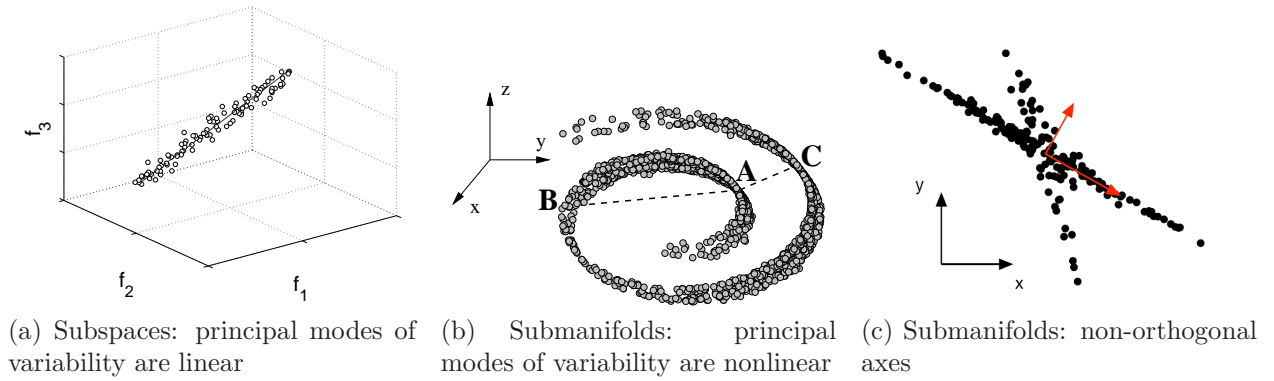


Figure 2.6: Addressing submanifolds: A problem for standard PCA (Figures 2.6b and 2.6c have been taken from [4] and edited).

### Kernel Based Nonlinear Objective Reduction

The method described in this section builds on classical (traditional linear) methods for dimensionality reduction such as PCA and multi-dimensional scaling (MDS) [139, 140]. The reformulation consists of using a kernel function [141], which is able to extract low-dimensionality structures by respecting some predefined constraints. The choice of the kernel is very important, since different kernels specialise in different types of low-dimensionality topological structures. The learning system embedded in this method needs to address two tasks in order to successfully unfold the data of any given unknown problem. The first task is to choose a kernel from a family of kernels, then select features that exist in the feature space of the selected kernel. One option is to employ a method known as Corretropy PCA, introduced in [95], that utilises a pre-specified kernel function inside its expectation function given by

$$V(\mathbf{x}, \mathbf{y}) = E[\chi(\mathbf{x}, \mathbf{y})], \quad (2.12)$$

where  $\chi$  is the kernel function and  $\mathbf{x}$  and  $\mathbf{y}$  are two random variables. This corresponds to the strategy followed by C-PCA-NSGA-II where the authors have utilised polynomial and Gaussian kernels. However, in order to be successful in addressing nonlinear type of data, it is required to customise the choice of the kernel. A kernel perfectly customised for a given problem will determine the “underlying submanifolds” existing in the dataset and allow us

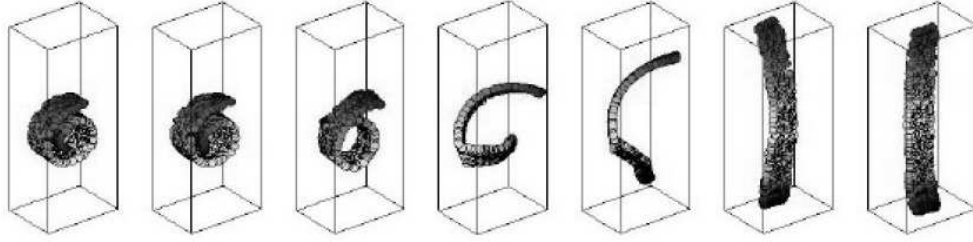


Figure 2.7: Maximum variance unfolding approach applied to the Swiss roll dataset showing different stages of the unfolding process (taken from [5]).

to separate the critical from redundant objectives. For this reason, instead of correntropy PCA the method here selected is Maximum Variance Unfolding (MVU). In the section that follows MVU is described in more detail.

### Maximum Variance Unfolding — The Procedure

Maximum variance unfolding (MVU) [5] is an unsupervised learning method designed to be applied to manifolds. The approach is able to construct (or learn) a kernel matrix by transforming the problem in a semidefinite programming instance. In order to build the kernel function, the method needs to respect local constraints, which preserve the data features. Such features are distances and angles between nearest data points, also called neighbours.

The authors that proposed the method have observed that “any slack in a piece of string serves to decrease the Euclidean distance between its two ends”, which lead them to the concept of MVU. In order to maximise the total sum of the pairwise distance between the dataset points, the concept of “unfolding” needs to be equated in an implicit manner in the kernel matrix to be learned. This unfolding strategy needs to preserve the angles and distances between data points in order for its neighbours to remain unchanged. As an example the transformation performed by MVU can be easily visualised in Figure 2.7, where a Swiss roll is unfolded in order to maximise the data global distance by maintaining the correct distances between neighbouring points.

Mathematically, first consider, high dimensionality inputs  $\{x_i\}_{i=1}^N$  and low dimensionality outputs  $\{y_i\}_{i=1}^N$ , where  $x_i \in \mathbb{R}^M$ ,  $y_i \in \mathbb{R}^m$ , and  $m \ll M$ . Let  $\eta_{ij} \in \{0, 1\}$  denote whether inputs  $x_i$  and  $x_j$  are  $k$ -nearest neighbours. Then, the transformation performed by MVU is described by the optimisation problem given by

$$\left. \begin{aligned} & \text{Maximise } \sum_{ij}^N \|y_i - y_j\|^2, \\ & \text{subject to constraints:} \\ & (1) \ \|y_i - y_j\|^2 = \|x_i - x_j\|^2, \forall (i, j) \text{ with } \eta_{ij} = 1 \text{ and} \\ & (2) \ \sum_i^N y_i = 0, \end{aligned} \right\} \quad (2.13)$$

where the first constraint ensures that distances between nearby inputs match distances between nearby outputs, and the second constraint guarantees a unique solution by centring the outputs on the origin. The above optimisation problem, can be reformulated as a semidefinite programming (SDP) problem [142] by defining the inner product matrix  $K_{ij} = y_i \cdot y_j$  as follows

$$\left. \begin{aligned} & \text{Maximise } \text{trace}(K) = \sum_{ij}^N \frac{(K_{ii} - 2K_{ij} + K_{jj})}{2N}, \\ & \text{subject to constraints:} \\ & (1) \ K_{ii} - 2K_{ij} + K_{jj} = G_{ii} - 2G_{ij} + G_{jj}, \forall \eta_{i,j} = 1, \\ & (2) \ \sum_{ij}^N K_{ij} = 0 \text{ and} \\ & (3) \ K \succeq 0, \text{ implying } K \text{ is positive-semidefinite,} \end{aligned} \right\} \quad (2.14)$$

where  $G_{ij} = x_i \cdot x_j$  and  $x_i \in \mathbb{R}^M$ . The learnt matrix  $K$  would be of size  $N \times N$  and the third constraint certifies positive-semidefiniteness of  $K$  which helps to interpret the kernel matrix as storing the inner products of vectors in a space with any finite or infinite number of dimensions (Hilbert space).

### Maximum Variance Unfolding — For Feature Selection

In the previous section, the general approach of MVU was provided, while here a small modification to the way MVU perceives input data and how mathematically that will help in learning the desired kernel is described.

A novel way of using MVU was proposed in [7] and named MVU-PCA-NSGA-II. The method was created in order to overcome the limitations of PCA-NSGA-II. To start, MVU treats a time sample (or a solution with a certain dimension) as an individual input. In EMO, a population is represented by its  $N$  individuals with  $M$  number of objectives, which are considered MVU inputs. Consequently, the learned matrix by MVU would be of size  $N \times N$ , which is incompatible with a desired  $M \times M$  kernel matrix that would correlate the importance of each objective. In order to address this dimensionality issue, the authors treat each “objective function” individually as a separate input to MVU. Hence, the philosophy of MVU would still hold since the data is been treated as a vector. This allows to correlate the importance of objectives into directions of large variance extracted from eigen-decomposition of the learned kernel matrix, in order to “unfold” a manifold. So then, the input variables could be treated as input features  $\{x_i\}_{i=1}^M$  and a  $M \times M$  kernel matrix could be retrieved. Considering the above, the formulation of the SDP problem is given by

$$\left. \begin{aligned} &\text{Maximise } \text{trace}(K) = \sum_{ij}^M \frac{(K_{ii} - 2K_{ij} + K_{jj})}{2M}, \\ &\text{subject to constraints:} \\ &\quad (1) \ K_{ii} - 2K_{ij} + K_{jj} = \dot{G}_{ii} - 2\dot{G}_{ij} + \dot{G}_{jj}, \forall \ \eta_{i,j} = 1, \\ &\quad (2) \ \sum_{ij}^M K_{ij} = 0 \text{ and} \\ &\quad (3) \ K \succeq 0, \text{ implying } K \text{ is positive-semidefinite,} \end{aligned} \right\} \quad (2.15)$$

where  $\dot{G}_{ij} = x_i \cdot x_j$  and  $x_i \in \mathbb{R}^N$ . As in Equation 2.14, the first constraint ensures that



distances between nearby inputs match distances between nearby outputs, the second constraint guarantees a unique solution by centring the outputs on the origin and the third constraint certifies positive-semidefiniteness of  $K$ . Since the size of the learnt matrix  $K$  is  $M \times M$ , the importance of each objective can now be correlated which is explored by MVU-PCA-NSGA-II. The procedure followed by MVU-PCA-NSGA-II is the same as PCA-NSGA-II (as described in Section 2.3.2) and the only difference is the replacement of the correlation matrix  $R$  (given in Equation 2.11) by the  $K$  matrix (give in Equation 2.15).

### 2.3.3 Online Implementation of the Decision Support

When an objective reduction algorithm operates *after* an MOEA run (*offline* reduction) the aim is to provide aid and support to the DM before the decision-making process. However, when objective reduction is implemented *during* the MOEA run (*online* reduction) the aim is to improve the computation efficiency (less number of function evaluations) and search ability (better POF-approximation) of an MOEA. Considering this, the currently available *online* objective reduction algorithms are as follows:

I The authors in [2] have explored the benefits of integrating DRP [92] into SIBEA [40] in tackling MaOPs. Towards this, the authors have modified SIBEA by applying objective reduction every  $G$  generations, where  $G$  is user defined, to decide which objectives are chosen for optimisation and which ones are neglected during the next  $G$  generations. While the original algorithm is referred to by the authors as  $\text{SIBEA}_{ref}$ , two versions have been introduced which are as follows:

- (a)  $\text{SIBEA}_{random}$ : selects a subset of objective with size  $k$  every  $G$  generations randomly, where  $k$  is given in advance.
- (b)  $\text{SIBEA}_{online}$ : the greedy algorithm for  $k$ -EMOSS, given in [92], is applied on the current population to decide which objectives are considered in the next  $G$  generations.

For validation the authors have utilised a modified version of the DTLZ2 [27] problem, known as DTLZ2<sub>BZ</sub>. Two versions of the DTLZ2<sub>BZ</sub> test problem have been considered: one where the objectives are scaled and another where the objectives are unscaled. Moreover, the experimental results have shown that: (i) for both versions of DTLZ2<sub>BZ</sub>, SIBEA<sub>random</sub> was the worst performing algorithm; (ii) for scaled DTLZ2<sub>BZ</sub>, SIBEA<sub>online</sub> was the best performing algorithm; and (iii) for unscaled DTLZ2<sub>BZ</sub>, SIBEA<sub>ref</sub> was the best performing algorithm except for the nine-objective problems with  $k = 3$  and  $k = 4$ .

II In [96] the authors have explored whether the integration of UFS [3] into NSGA-II can be beneficial in tackling MaOPs. Towards this, the authors have proposed two different schemes which are as follows:

- (a) REDGA-S: reduces periodically the number of objectives during the search until the required objective subset size has been reached, and by the end of the search the original objective set is used again.
- (b) REDGA-X: alternates between the reduced set and the entire objective set during the search process.

For validation the authors have used DTLZ2<sub>BZ</sub> and Knapsack [143] test problems scaled up to ten objectives. The obtained results have shown that both reduction schemes outperform NSGA-II and besides reducing the execution time of an MOEA it can also lessen the limitations associated with Pareto-dominance when dealing with MaOPs. However, it is highlighted that for both schemes the parameters have to be carefully selected.

The previous mentioned *online* objective reduction algorithms have been implemented as an iterative procedure, where each iteration is comprised of an MOEA run and an objective reduction algorithm application. The iterative procedure is advantageous over a non-iterative one since as the number of objectives is reduced gradually, it is expected for the quality of the POF-approximation/representation to improve. Given that most MOEAs provide a poor

POF-approximation when dealing with MaOPs it is desirable for an objective reduction approach to be implemented as an iterative procedure.

Moreover, another important aspect is the selected test problems used to validate the above objective reduction algorithms. Notably, the authors have selected the non-redundant DTLZ2<sub>BZ</sub> and Knapsack test problems. An  $M$ -objective problem is considered to be non-redundant if the reduced number of objectives is equal to  $M$ , i.e.,  $m = M$ . For this type of problems all objectives are essential to describe the POF. Therefore, it is expected some information loss if an objective is omitted for problems that belong to this category. Given this, the benefits of *online* objective reduction have not been explored for those problems where the reduce number of objectives is less than  $M$ , i.e.,  $m < M$ .

### 2.3.4 Timing the Decision Support

One of the most important parameters that one needs to specify before running an MOEA is the number of generations. The number of generation specifies for how long an MOEA is allowed to run until the population is declared as final. When the number of generations is set appropriately, the population travels across the search space until it converges, or until it stagnates. Since the relationships between the objectives vary from problem to problem it is common for different problems to require different number of generations until stability is achieved. Therefore, selecting the number of generations appropriately for a given problem is not a trivial task. For instance, if the number of generations is too low, the population might be still far away from the POF. On the other hand, if the number of generations is much higher than the number of generations it takes for the population to stabilise, computational resources will be wasted.

Given the above background, it becomes even more crucial to select the number of generations appropriately when dealing with MaOPs due to the following reasons:

1. If the number of generations is too high the waste of computational resources might increase exponentially with an increase in the number of objectives.

2. When dealing with MaOPs since most MOEAs poorly approximate the POF, it is impractical to rely on a pre-specified level of convergence as a criterion to cease the run.

Moreover, to address the fundamental question as to when the decision support should be applied during an MOEA run, one possibility is to rely on a termination criterion capable of determining stability in multi-objective context. Knowing that during the search for the Pareto-optimal solutions the population changes as it travels in the search space. It implies that the correlation-structure and as well the dominance relations of the population might change as well. Therefore, for the decision support analysis to be the most accurate and reliable as possible, it is important to apply the decision support to a stable population.

### Available Approaches

Termination criterion in multi-objective optimisation is a recent area of research. The available termination criterion algorithms utilise quality indicator as heuristics to measure differences between populations, which can be thought of as a kind of distance measure. Given this, the existing termination criterion algorithms for multi-objective optimisation are as follows:

1. In [97] an online stopping criterion algorithm, namely steady performance stopping criterion, is proposed and integrated into NSGA-II. The algorithm uses the maximal crowding distance computed for a defined number of generations and considers that stability is achieved when the variation of the maximal crowding distance mean is below an user-defined threshold. It is known that, for a population with  $N$  solutions and  $M$  objectives, the worst-case complexity of crowding distance assignment is  $O(M(2N) \log(2N))$ . This means that the proposed algorithm is very computationally demanding unless the crowding distance is already provided by the MOEA, as happens for NSGA-II. For validation the authors have utilised four two-objective ZDT test problems.

2. A more general stopping criterion algorithm that can be applied to other MOEAs, named by the authors as Marti–Garcia–Berlanga–Molina (MGBM) criterion, has been proposed in [98]. It is based on a proposed performance indicator that quantifies the ratio of dominated solutions in the population, named as mutual dominance indicator. The algorithm stops the run when the ratio between two consecutive generations is considered stable for that purpose using a modified Kalman filter and a defined threshold. The authors have integrated the termination criterion into NSGA-II and SPEA2 algorithms. For validation the following versions of the DTLZ test problems have been considered: (i) three- and ten-objective DTLZ3 and (ii) three-objective DTLZ7. In an extended work [99], besides the above, the authors have also integrated the termination criterion into PESA algorithm and all test problems have been scaled up to ten-objectives. Moreover, DTLZ6 test problem has also been considered. Later, in [144] the mutual dominance indicator is replaced by the normalised objective values variance and the termination criterion is named as fitness homogeneity indicator (FHI). The termination criterion has been integrated into NSGA-II, SPEA2 and HypE algorithms. For validation the test problems DTLZ3 and DTLZ6 scaled up to ten objectives have been utilised.
3. In [145] a set of performance indicators are applied to a series of MOEA runs and the criterion is named as offline convergence detection (OFCD). The performance indicators utilised are: generational distance, hypervolume, and the spread metric. After several runs are conducted, the Kolmogorov-Smirnov statistical test is used to detect convergence. It is assumed that the population has converged if there is no significant change in the collected performance indicators along the generations. The algorithm is theoretically promising but it cannot be applied *online* (during an MOEA run) since it requires a set of MOEA runs and it is also considered to be heavily parametrised. In a work [146] that followed, the authors proposed an *online* implementation of a similar algorithm and named it online convergence detection (OCD). Towards that, the same

performance indicators are considered and the Kolmogorov-Smirnov statistical test is replaced by the  $\chi^2$ -variance test and  $t$ -test. The criterion stops the simulation when both tests detect convergence or stagnation. Recently, the previously mentioned *offline* and *online* algorithms have been compared in [101]. The termination criterion algorithms have been applied to NSGA-II and SMS-EMOA populations. For validation the test problems considered have been the three-objective DTLZ2 and the two-objective ZDT1, ZDT2, ZDT4 and FON [147].

4. In another work [100] the performance indicators hypervolume,  $\epsilon$ -indicator, and mutual dominance rate are combined during the analyses of the convergence. The algorithm, named as least squares stopping criterion (LSSC), computes the adjustment of the indicator values to an uniform model using for that the residue captured by a least squares approximation. As a complement, the algorithm also uses the slope of the model to muster confidence on the criterion accuracy. The termination criterion algorithm has been integrated into NSGA-II, SPEA2 and PESA algorithms. For validation the three-objective DTLZ3, DTLZ6 and DTLZ7 have been utilised.
5. In [148], a non-dominance-based indicator that quantifies the number of non-dominated solutions by comparing two populations from different generations has been proposed. The algorithm is then combined with an utility function which is used to determine when termination should take place. The termination criterion is integrated into NSGA-II and for validation it is applied to two analytical problems, namely TNK [149] and OSY [150], and to two real-world crashworthiness problems, namely knee bolster design and multi-disciplinary optimisation.
6. Recently, in [151] a proximity measure based on Karush–Kuhn–Tucker (KKT) conditions [152] is used to test the convergence of a set of non-dominated solutions to the POF. The concept has been applied to single objective problems in an earlier study found in [153] and a more comprehensive version can be found in [154]. The new pub-

lication has extended the application to multi-objective optimisation and the authors endeavour to apply the concept as a termination criterion for MOEAs. For validation, the authors have selected two test problems and NSGA-II to generate the population. One is the two-objective ZDT1 and the other is the three-objective version of DTLZ5. The experimental results have revealed that the proposed measure has a smoothly reducing property and converges to zero in most cases, however, for harder problems it requires a local search effort.

As suggested in [144], a termination criterion consists of two processes namely: (i) progress indicator, and (ii) evidence gathering. The former applies a set of heuristics (or distance measures) at each generation during the optimisation process. The latter collects the information generated by the progress indicator along generations and employs a criterion for determining when the algorithm should cease to run. Considering the two mentioned processes, a summary of the previously described termination criterion algorithms is represented in Table 2.1.

### **Difficulties in Many-objective Optimisation**

The termination criterion algorithms often utilise quality indicator as heuristics to measure differences between populations, which can be thought as a kind of distance measure in multi-objective context. Distance measures are commonly used to compare the performance of different MOEAs [155, 156] and some examples are: hypervolume [36],  $\epsilon$ -indicator [156], generational distance [121], crowding distance [31], mutual dominance rate [98] and population diversity [157]. These distance measures operate on the objective vectors of the non-dominated solution sets and attempt to measure the quality of a set (with or without a reference) using a single number. Notably, here the aim is to detect changes between two consecutive populations. Therefore, the quality needs to be measured between two non-dominated solution sets. However, when dealing with MaOPs the following difficulties arise:

Table 2.1: Selected multi-objective optimisation stopping criterion algorithms. The algorithms are organised by order of publication where the bottom elements are the most recent publications.

Algorithm Name	Progress Indicator	Evidence Gathering
Steady performance stopping criterion [97]	Maximal crowding distance	Variance of the maximal crowding distance
MGBM [98, 99]	Mutual domination rate	Modified Kalman filter
OFCD [145, 101]	Generational distance, spread and hypervolume	Kolmogorov-Smirnov test
OCD [146, 101]	Generational distance, spread and hypervolume	$\chi^2$ -variance test and $t$ -test
FHI [144]	Variance of the normalised objective values	Modified Kalman filter
LSSC [100]	Hypervolume, $\epsilon$ -indicator and mutual domination rate	Least squares approximation of an uniform model and respective slope
Non-dominance-based stopping criterion [148]	online Consolidation ratio convergence indicator	Utility function-based approach
Proximity measure based KKT [151]	KKT-proximity measure	—

- I Non-domination based distance measures: these distance measures rely on non-domination to rank the solutions. It is known that, when the number of objectives increases, the number of non-dominated solutions as well [63]. As a result, when the number of objectives increases these type of distance measures become ineffective in detecting changes. Examples are  $\epsilon$ -indicator and mutual dominance rate.
- II High computational cost: distance measures whose computational cost increases exponentially with an increase in the number of objectives are for instance: hypervolume, generational distance and crowding distance. Due to the high computational cost, the application of these distance measures to MaOPs is very difficult due to practical reasons.
- III Gaussian distribution assumption: The distance measures that take the average of all individual distances make a stringent assumption that the individual distances need to follow a Gaussian distribution. This makes them inappropriate to detect changes



between two consecutive populations when the number of objectives increases. One example is generational distance.

The previous three Items calls for a new distance measure capable of detecting changes between two consecutive high-dimensional non-dominated solution sets with the following three desirable features:

1. The distance measure cannot be based on non-domination operator. Otherwise, it will be ineffective in detecting changes when the number of objectives increases.
2. When the number of objectives increases the computational complexity needs to be linear or relatively low. Otherwise, due to practical reason it cannot be applied to MaOPs.
3. The distance measure should not make any assumptions regarding the distribution of solutions that are being evaluated. Otherwise, it might not be flexible enough to capture the distribution of solutions.

The success of a termination criterion is highly dependent on the utilised distance measure(s). However, many distance measures fail to detect changes on a population when the number of objectives increases. Here, it is envisaged that if a distance measure possesses the three features mentioned above a termination criterion should be able to detect changes in high-dimensional space. This is important when dealing with MaOPs where  $M \gg 4$ .

## 2.4 Research Gaps

The research gaps identified in the literature are described in three separate sections. Each section corresponds to a previous literature review section as follows:

1. In Section 2.4.1 the identified research gaps focus on the available objective reduction approaches (mentioned in Sections 2.3.1 and 2.3.2). In particular it will be discussed

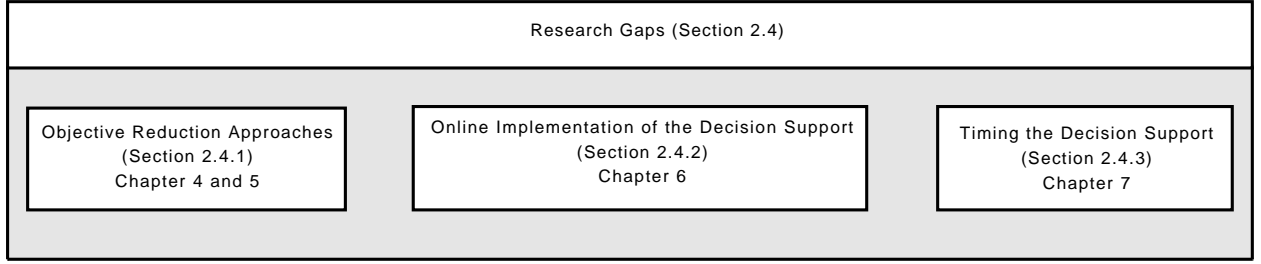


Figure 2.8: Structure of the research gaps section.

with respect to objective reduction approaches: (i) the criterion employed, (ii) the inability in dealing with *noisy* solutions, and (iii) the sensitivity of the approaches when applied to test problems with different degrees of redundancy. The research gaps identified in this section will be addressed in Chapters 4 and 5.

2. In Section 2.4.2 the identified research gaps focus on the existing *online* objective reduction algorithms (mentioned in Section 2.3.3). The research gaps identified in this section will be addressed in Chapter 6.
3. In Section 2.4.3 the identified research gaps focus on the existing termination criterion algorithms for multi-objective optimisation (mentioned in Section 2.3.4). The research gaps identified in this section will be addressed in Chapter 7.

Moreover, to facilitate the visualisation of this section organisation, a diagram is provided in Figure 2.8.

### 2.4.1 Objective Reduction Approaches

Different objective reduction algorithms have been proposed in the literature as shown in Section 2.3. Although these different approaches have given an important contribution to this research line, the following questions remain unaddressed.

1. While it is known that most existing MOEAs fail to provide a good POF-approximation (in terms of convergence and diversity), it is not known what criterion ought to be used so that more accurate and reliable inferences about the dimensionality and composition

about the true POF can be made. For example, it needs to be studied whether preservation of the dominance relations (as is the case with the DRP-based algorithms) of the given solution set is a good criterion or whether the preservation of the correlation-structure (as is the case with PCA based algorithms) is a better criterion, in the above context.

2. If the difference in the dominance relations or the correlation-structure of the solution set obtained by an MOEA and the true POF be termed as *noise*, then accurate inferences about the latter based on the former, would require some kind of *de-noising* operation. Such *de-noising* operation should ideally be problem-specific and should adapt itself based on the problem in consideration. No existing research discusses the need for and the mode of *de-noising* when dealing with *noisy* solution sets.
3. If the dimension of the POF is less than the number of objectives in the problem, i.e.,  $m < M$ , the problem can be referred to as redundant, and if  $m = M$ , the problem can be referred to as non-redundant. In the existing literature, the application of the DRP-based algorithms has largely been on non-redundant problems, while the PCA based algorithms have largely focussed on redundant problems. It is interesting to note that none of these algorithms have been extensively tested for both the redundant and non-redundant problems, implying that the robustness of both these sets of algorithms is not established. In this background, it is not known if a particular approach/algorithm is structurally limited to only a specific (redundant or non-redundant) class of problems, or are these generalised enough to tackle both class of problems.
4. The concepts known as  $\delta$ -MOSS and  $k$ -EMOSS are defined as the  $\delta$ -minimum objective subset and the minimum objective subset of size  $k$  with minimum error, respectively. The former imitates a situation where the DM allows for a  $\delta$  ( $0 \leq \delta \leq 1$ ) error, and wants to know the smallest objective set-retaining which would ensure that the error incurred is just less than or equal to  $\delta$ . The latter imitates a situation where the

DM specifies a fraction  $p$  ( $0 \leq p \leq 1$ ) of the original number of objectives ( $M$ ) to be retained, and wants to know an objective set of size  $k = \lceil pM \rceil$  that corresponds to the minimal error. These two concepts have been explored by the DRP-based algorithms in [2] and also by the UFS in [3]. However, the PCA based based objective reduction algorithms have not yet explored these two concepts.

### 2.4.2 *Online* Implementation of the Decision Support

Depending at which stage an objective reduction algorithm is applied during the optimisation process the implementation can be considered: (i) *offline reduction* if the objective reduction algorithm is applied after an MOEA run and (ii) *online reduction* if the objective reduction algorithm is applied during an MOEA run. Given this, the relevance of the *offline* or *online* objective reduction needs to be seen in the light of the following:

I The possible scenarios for the dimensionality of the true POF: The case of an  $m$ -dimensional POF for an  $M$ -objective problem could broadly belong to one of the following categories:

- (a)  $m = M$ : Here, objective reduction while revealing that the problem can not be reduced, could inform the DM(s) of the relative importance of all the objectives.
- (b)  $m < M$  and  $m \geq 4$ : Here, objective reduction besides simplifying the problem by eliminating the redundant objectives, could also inform the DM(s) of the relative importance of the *essential* objectives.
- (c)  $m < M$  and  $m \leq 3$ : Here, objective reduction by virtue of reducing the number of objectives in a given problem to three or less, can facilitate its solution by any of the existing MOEAs.

Notably, most existing MOEAs are likely to fail in approximating the complete POF for the problems considered in Item I(a) and I(b), as the dimension of the POF is more than three. In these cases only a part of the POF could be obtained through

the intervention of the DM, and hence the decision support offered by *offline* objective reduction is appropriate. Contrary to these cases, the case in Item I(c) provides ground for integration of the objective reduction procedure into the MOEA search, such that the complete POF of the problem could be retained at the end of the MOEA run. In other words, the benefits of *online* objective reduction could be most realised in problems where regardless of the number of objectives, the POF is two- or three-dimensional.

- II The quality of the solution set on which objective reduction analysis is performed: It has previously been highlighted that most existing MOEAs poorly approximate the POF for MaOPs. In such situations, the inferences drawn by the objective reduction analysis about the dimensionality and composition of the POF may be erroneous. Given this, it is desirable that objective reduction analysis is performed iteratively, where an objective identified as redundant in a particular iteration is not totally eliminated and is allowed an opportunity for being reconsidered in a subsequent iteration. Besides accounting for the quality of the underlying solution set, an iterative scheme will also be able to account for problems where the relationships (conflict or non-conflict) between the objectives may vary across the search space, in that, an objective inferred as redundant in the initial iterations may actually be *essential* with regards to the true POF, and vice-versa.

Given the above and in terms of the existing literature, the following research gaps could be observed:

1. The *online* implementation of the PCA based algorithms [6, 7] has not been explored by their authors. The PCA based algorithms have only explored the *offline* implementation.
2. The *online* implementation of the algorithms based on DRP [2] and UFS [3] has been explored, but the focus has been only on one problem belonging to the category where

$m = M$ , mentioned in Item I(a). Ironically, the problems in Item I(b) and I(c), where *online* objective reduction is more relevant, have been ignored.

3. In an *online* implementation no attempt has been made in the literature to demonstrate that if an essential objective is erroneously eliminated as redundant at some stage, it could be re-introduced at a subsequent stage, and then be correctly identified as essential.

### 2.4.3 *Timing* the Decision Support

Different termination criterion algorithms for multi-objective optimisation exist in the literature, as mentioned in Section 2.3.4. These approaches have been important in addressing the fundamental question as to when an MOEA should cease to run and declare the population as final. However, consider the following drawbacks of the existing approaches:

- I The existing termination criterion algorithms rely on distance measure(s) to evaluate changes in a set of consecutive populations. Some difficulties (identified in Section 2.3.4) have prevented the existing distance measures in dealing with MaOPs. In resume the identified difficulties are: (i) the distance measures that rely on non-domination are not able to differentiate between solutions when the number of objectives increases; (ii) some distance measures computational complexity increases exponentially with an increase in the number of objectives; and (iii) distance measures that take the average of all individual distances make a stringent assumption that individual distances have to follow a Gaussian distribution.
- II The majority of the available termination criterion publications have concentrated on test problems with two objectives. The publications where MaOPs have been considered have utilised non-domination based distance measures [98, 99]. Hence, once all the population becomes non-dominated these termination criterion algorithms would suggest the run to be ceased. This might reveal to be incorrect if the utilised MOEA is not

based on non-domination. For instance, consider a population where all the solutions are non-dominated. It could happen that an MOEA not based on non-domination might still be able to improve the population. In this situation a termination criterion based on non-domination would not be able to detect any further changes in the population.

Given the above background, the following gaps have been identified in the literature:

1. The difficulties that have prevented existing distance measures in detecting changes when applied to high-dimensional datasets have not been addressed. This prevents the successful application of existing termination criterion algorithms to MaOPs.
2. No termination criterion has been extensively tested for a wide variety of test problems with small and large number of objectives. This implies that the robustness of these algorithms is not yet established in multi- and many-objective optimisation.

In regards to the objective reduction decision support the following remains unaddressed. During the validation of the existing objective reduction algorithms their respective authors have selected a number of generations arbitrarily. As a result, it could happen that the reported analysis might change if the MOEA would have been allowed to run for more generations. Therefore, to assure that the decision support analysis is the most accurate and reliable as possible, given the MOEA capabilities, it is important to specify the number of generations appropriately for the given problem. In this way, it is possible to take full advantage of the MOEA capabilities without wasting computational resources unnecessarily, assuming that the termination criterion computational complexity is not too high. Given this, the application of a termination criterion to enhance the accuracy of a decision support has not been explored by the available objective reduction approaches.

## Chapter 3

# Research Aim and Objectives

The aim and objectives of this thesis work are identified in this chapter. To achieve these, the methodology that has been adopted and the corresponding scope of the research are also described. This chapter concludes with the main contribution and key findings of this thesis. The structure of this chapter is as follows:

1. research aim and objectives (Section 3.1);
2. methodology (Section 3.2);
3. scope of the research (Section 3.3);
4. main contribution and key findings (Section 3.4).

### 3.1 Research Aim and Objectives

The aim of this thesis is to develop a framework that offers both *offline* and *online* decision support<sup>1</sup>, enabling the articulation of DMs preferences with *objectivity* (a rational basis), *repeatability* (identical preferences for identical options), *consistency* (alike preferences across

---

<sup>1</sup>The terms *offline* and *online* refer to the type of objective reduction implementation. In that, an *offline* implementation run applies the objective reduction approach *after* the MOEA run while the *online* applies the objective reduction approach *during* the MOEA run.



multiple interaction stages), and *coherence* (alike preferences by multiple DMs). To achieve this aim, this thesis has the following objectives:

1. identification of a meaningful criterion for developing a decision support: the developed decision support needs to be capable of dealing with *noise* in the input data, i.e., to handle solutions whose dominance relations are different from those characterising the POF.
2. development of a machine learning based *offline* objective reduction framework towards the envisioned decision support with the following characteristics:
  - (a) that could *learn* the problem–structure of a given problem by identifying—the smallest set of conflicting objectives which can generate the same POF as that of the original problem.
  - (b) that could identify the smallest objective sets corresponding to an error specified by the DM.
  - (c) that could identify objective sets of size specified by the decision maker, that corresponds to a minimum error.
3. to explore the benefits of implementing *online* the proposed framework: the proposed framework follows an *offline* implementation which means that an MOEA is run for a pre-defined number of generations and the framework is only applied to the final population. In an *online* implementation the framework is applied during the MOEA run which means that it can be applied to an earlier population other than the final.
4. an address of the fundamental question as to when the proposed framework should be applied: as the population significantly keeps changing along an MOEA run, so may its POF–approximation/representation. To guarantee timely results with the most accurate decision support analysis it is important to determine when the MOEA population stagnates (implicitly its POF–approximation/representation too).

5. experimental validation and comparative analysis through a wide range of test problems and real-world problems.

## 3.2 Methodology

To realise the above aim and objectives, the following methodology (summarised in Figure 3.1) has been adopted:

- (a) literature review: this includes a review of the state-of-the-art in many-objective optimisation, with a focus on objective reduction approaches. The literature review has been realised in Chapter 2.
- (b) identification of many-objective test suites: this includes identification of test problems which are scalable to many-objectives, and also real-world problems available in the public domain. Each  $M$ -objective problem is characterised by an original objective set denoted by  $\mathcal{F}_0 = \{f_1, f_2, \dots, f_M\}$  and an *essential* objective set—the smallest set of conflicting objectives ( $\mathcal{F}_{\mathcal{T}}, |\mathcal{F}_{\mathcal{T}}| = m \ (m \leq M)$ ) which can generate the same POF as that obtained by  $\mathcal{F}_0$ . Then, from an objective reduction perspective, the problems can be categorised as *highly-redundant*, *moderately-redundant*, and *non-redundant*, depending on the difference between  $m$  and  $M$ . The identification of many-objective test suites has been conducted along Chapters 4–7.
- (c) development of a robust machine learning based *offline* objective reduction framework: this includes investigation of the strength and limitations of existing objective reduction algorithms based on principal component analysis and maximum variance unfolding. The existing algorithms are generalised to deal with *noise* in the input data, towards investigating MaOPs, with varying degree of redundancy, from the following perspectives:

I revelation of an *essential* objective set: here, the aim is to determine an *essential* objective set  $\mathcal{F}_{\mathcal{T}}$ , as defined above.

- II preference-ranking of all the objectives: here, the aim is to obtain the preference-weight ( $w_i$ ) for each objective, such that  $w_i \geq 0$  and  $\sum_{i=1}^M w_i = 1$ .
- III  $\delta$ -MOSS perspective: here, the aim is to determine the smallest objective (sub)set corresponding to an *a priori*-specified  $\delta$  ( $0 \leq \delta \leq 1$ ) error. Such a (sub)set is referred to as the  $\delta$ -minimal set, denoted by  $\mathcal{F}_{\{\delta\}s}$ , where  $|\mathcal{F}_{\{\delta\}s}|$  denotes its size. This analysis, referred to as  $\delta$ -MOSS (minimum objective subset) analysis, imitates a situation where the DM may allow for some  $\delta$  error and is interested in knowing the smallest objective (sub)set-retaining which ensures that the error incurred is *just less than or equal to*  $\delta$ .
- IV  $k$ -EMOSS perspective: here, the aim is to determine the objective (sub)set of *a priori*-specified size  $k$  that corresponds to minimal error. Such a subset is referred to as the  $k$ -minimal set, denoted by  $\mathcal{F}_{\{k\}s}$ , where the corresponding  $k$ -minimal error is denoted by  $\mathcal{E}_k^n$ . This analysis, referred to as  $k$ -EMOSS (minimum objective subset of size  $k$  with minimum error) analysis, imitates a situation where the DM may specify the fraction  $p$  ( $0 \leq p \leq 1$ ) of the original number of objectives ( $M$ ) that he/she is keen to retain, and, is interested in an objective (sub)set of size  $k = \lceil pM \rceil$  that corresponds to minimal error.
- V visual representation: Here, the aim is to develop a simple yet meaningful visual representation of the above analysis, that could serve as a snap-shot guide for the DMs to base their preferences on.
- (d) development of both *offline* and *online* implementations of the proposed framework. The *offline* implementation of the proposed framework addresses the first and second objectives and are realised in Chapters 4 and 5, respectively. The *online* implementation of the framework addresses the third objective and it is realised in Chapter 6.
- (e) development of an approach capable to identify when is more significant to apply the decision support framework during an optimisation process. The development of this

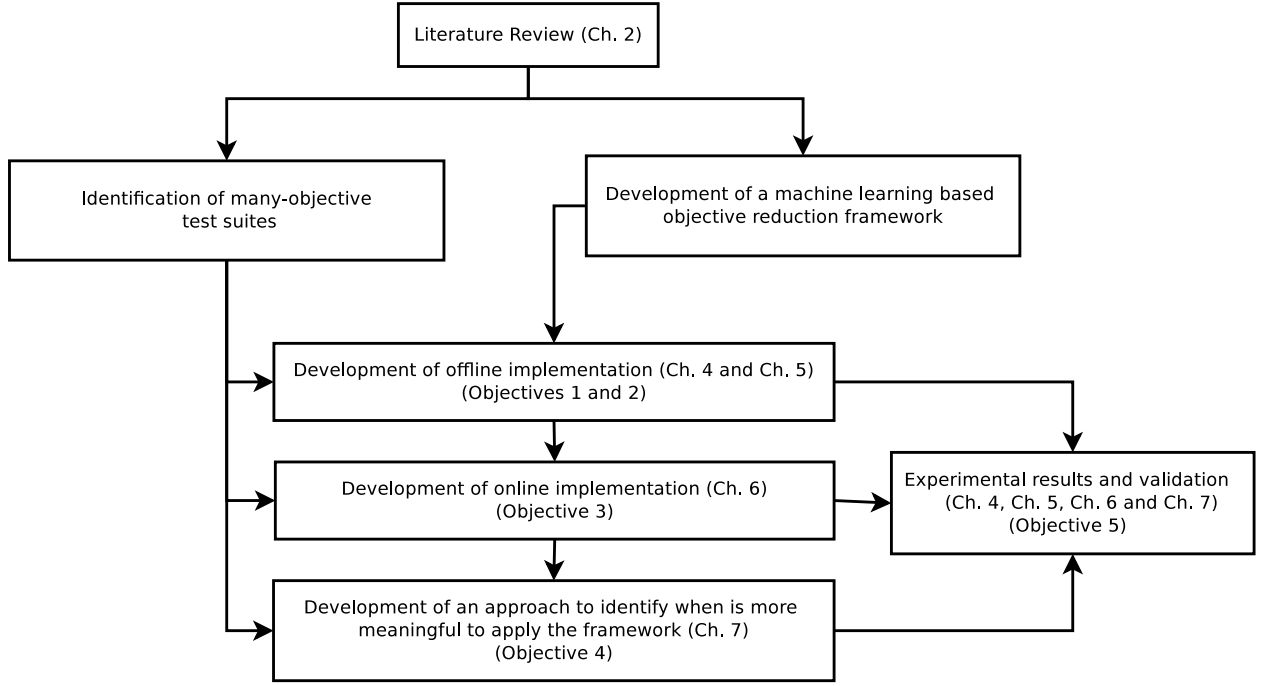


Figure 3.1: Steps of the methodology adopted by this thesis and its connection with proposed objectives and corresponding chapters.

approach addresses the fourth objective and it is realised in Chapter 7.

- (f) experimental results and their validation: this includes presentation of the experimental results based on the proposed approaches on a wide range of test and real-world problems and comparison of these results with those obtained from alternative approaches reported in literature. This addresses the fifth objective and it is realised along Chapters 4–7.

The decision support framework consists of several parts that have been addressed by separate objectives. This includes: (i) the decision support selected criterion (objective 1), (ii) the *offline* implementation (objective 2), (iii) the *online* implementation (objective 3) and (iv) the *timing* of the decision support (objective 4). Each one of these parts has its own literature, research gaps, proposed approach, experimental results, findings and contributions. To provide to the reader a chance to examine the findings/contributions of this thesis in the light of the identified research gaps and link these to the corresponding literature, each aforementioned objective has its own separate chapter as mentioned in Section 3.1.

### 3.3 Scope of the Research

The focus in this thesis is on evolutionary many-objective optimisation. Given that most existing MOEAs are known to fail in providing a good POF-approximation for MaOPs, their applicability on such problems is questionable at first place. In this background, it is meaningful to identify:

- (i) if all the objectives in a given problem are *essential* to describe the complete POF or not. Based on such an identification, if the true POF of an  $M$ -objective ( $M \gg 4$ ) problem could be described by just  $m$  objectives ( $m \leq M$ ), an otherwise unsolvable problem (by an MOEA) may become solvable (for instance, if  $m \leq 3$ ).
- (ii) the error that may be incurred if some of the *essential* objectives are to be eliminated for the sake of problem simplification. This information may help the DMs in articulating their preferences with a more *rational* basis.

Recognition of the above explains the aim of this thesis which is to develop a framework that offers both *offline* and *online* decision support, enabling the articulation of DMs preferences with *objectivity*, *repeatability*, *consistency*, and *coherence*.

In the wake of the aim of this thesis, its scope encompasses both the unconstrained and constrained MaOPs (both test problems and some real-world problems) with largely varying number of variables (both real and discrete) and objectives (varying from four to 50).

### 3.4 Main Contribution and Key Findings

The main contribution to knowledge of this thesis is the proposal of a machine learning based framework that offers both offline and online decision support, facilitating objectivity, repeatability, consistency and coherence in the DMs' preferences, when dealing with MaOPs. The key findings of this thesis are:

1. In Chapter 4, it has been shown that the DRP based algorithms are extremely sensitive to *noise* and therefore these algorithms fail to accurately identify an *essential* objective set for noisy data sets. This implies that the merit of preserving the dominance relations of noisy data sets is itself questionable. The proposed L-PCA and NL-MVU-PCA have reported a significantly better performance when compared with DRP based algorithms. This can be attributed to a reasonably good POF-representation that characterises the majority of the data sets, and for those instances where the POF-representation was not good an *essential* objective set could still be obtained: (i) by the iterative approach of the framework and (ii) by the algorithms *de-noising* features. Moreover, NL-MVU-PCA outperforms L-PCA owing to the fact that the former applies PCA to the kernel space that allows for accurate determination of the principal components, which is not possible when PCA is applied directly to the objective space.
2. In Chapter 5, the DRP based exact algorithm when applied to unnoised signal it could accurately identify the  $\delta$ -minimal set for  $\delta = 0$  and  $k$ -minimal error for  $k = I$  in case of DTLZ5( $I, M$ ) and  $k = M$  in case of non-redundant problems. The same algorithm has failed to capture the correct problem features for  $\delta$ -MOSS and  $k$ -EMOSS analysis corresponding to  $\delta > 0$  and  $k < M$ , respectively. The application of NL-MVU-PCA to the same problems revealed that NL-MVU-PCA is able to capture the expected trend and that the failure of the DRP exact algorithm could be attributed to the stringent assumptions that the method is based on. The same results are observed when both algorithms are applied to noised signal where the failure of DRP based exact algorithm can be also attributed to its sensitivity to *noise*.
3. In Chapter 6, it was found that when the preference information of the decision support in a particular iteration was erroneous, the proposed *online* implementation that features a probabilistic retention of all objectives at a given instant is able to improve the correctness of the decision support in subsequent iterations of the framework. This

has been demonstrated in comparison with the *offline* implementation despite higher demands for computational resources.

4. In Chapter 7, it was found that the proposed *dissimilarity* measure is capable of identifying on its own when the underlying MOEA stabilises which implies that: (i) for two- and three-objective problems it identifies when the underlying MOEA converges to the POF and (ii) for MaOPs it identifies when the MOEA population has stagnated away from the POF. This provides an indication to the decision support as to when it is more meaningful to apply the machine learning based objective reduction analysis.

### 3.5 Summary

This chapter has presented the aim and objectives of this thesis work, the methodology adopted in order to achieve them, the scope of the research, and the main contribution and key findings of this thesis.

The aim is to develop a framework that offers both *offline* and *online* decision support, enabling the articulation of DMs preferences with *objectivity*, *repeatability*, *consistency* and *coherence*. The objectives are: (1) identification of a meaningful criterion for developing a decision support; (2) development of a machine learning based *offline* framework towards the envisioned decision support; (3) to explore the benefits of implementing *online* the proposed framework; (4) an addressal of the fundamental question as to when the proposed framework should be applied; and (5) experimental validation and comparative analysis through a wide range of test problems and real-world problems.

To realise the above objectives the methodology adopted consists of:

- (a) literature review (realised in Chapter 2).
- (b) identification of many-objective test suites (realised along Chapters 4–7).
- (c) development of a robust machine learning based objective reduction framework from

the following perspectives: (I) revelation of an *essential* objective set, (II) preference-ranking of all the objectives, (III)  $\delta$ -MOSS perspective, (IV)  $k$ -EMOSS perspective and (V) visual representation.

- (d) development of both *offline* and *online* implementations of the proposed framework. The *offline* implementation of the proposed framework addresses the first and second objectives and are realised in Chapters 4 and 5, respectively. The *online* implementation of the framework addresses the third objective and it is realised in Chapter 6.
- (e) development of an approach capable to identify when it is more significant to apply the decision support framework during the underlying MOEA run. The development of this approach addresses the fourth objective and it is realised in Chapter 7.
- (f) experimental results and their validation. This addresses the fifth objective and it is realised along Chapters 4–7.

The scope of this thesis research work lies on evolutionary many-objective optimisation. More specifically, it builds on the recognition that the majority of MOEAs fail to provide a good POF-approximation for MaOPs, and therefore their applicability to such problems is questionable at first place. This recognition establishes that it is meaningful to identify: (i) if all the objectives in a given problem are *essential* to describe the complete POF or not; and (ii) the error that may be incurred if some of the *essential* objectives are to be eliminated for the sake of problem simplification. The identification of these two points can lead to: (a) for an  $M$ -objective problem with  $M \gg 4$ , if the problem could be described by just  $m$  objectives  $m < M$ , an otherwise unsolvable problem (by an MOEA) may become solvable (for instance, if  $m \leq 3$ ; and (b) the information derived from the problem simplification may help a DM in articulating his/her preferences with a more *rational* basis. Finally, the recognition of the last two points explains the aim of this thesis which is to develop a framework that offers both *offline* and *online* decision support, enabling the articulation of DMs preferences with *objectivity, repeatability, consistency, and coherence*.



## Chapter 4

# Machine Learning Based *Offline* Objective Reduction Framework

Two promising objective reduction approaches, namely dominance relation preservation (DRP) and machine learning based objective reduction, capture the preference-structure of an MOP by preserving the dominance relations or the correlation-structure of the corresponding non-dominated set, respectively. Knowing that the majority of the MOEAs provide a poor POF-approximation, when applied to MaOPs, it is important to determine which criterion can provide the most accurate and reliable inferences about the dimensionality and composition about the true POF. Due to the stochastic nature of the MOEAs, the non-dominated solutions obtained are approximations of the Pareto-optimal solutions. Therefore, the dominance relations characterising the obtained solutions may be different from those characterising the POF, a feature that is referred to as *noise*. For the existing objective reduction approaches to be able to handle *noisy* solution sets with accuracy, it is important to explore some kind of *de-noising* operation. Moreover, a test problem is denoted as *non-redundant* if the number of essential objectives needed to describe the POF is equal to the number of objectives in the problem, i.e.,  $|\mathcal{F}_{\mathcal{T}}| = M$ . On the other hand, if number of essential objectives is less than the number of objectives in the problem, i.e.,  $|\mathcal{F}_{\mathcal{T}}| < M$ ,

the problem is denoted as *redundant*. The redundancy in a problem is determined by the difference between the number of essential objectives and the number of objectives in the problem. That is, the higher that difference is, the more redundancy there is in the problem, and vice versa. Since test problems with different degrees of redundancy are available in the literature, it is important to explore the robustness of the existing objective reduction approaches in handling such problems.

In this background, this chapter proposes a machine learning based framework for linear and nonlinear objective reduction algorithms, namely L-PCA and NL-MVU-PCA. To interpret the conflict captured between objectives the former is based on principal components analysis (PCA) while the latter is based on maximum variance unfolding (MVU). The criterion adopted by the proposed framework preserves the correlation-structure of the given solution set, and is found to be more accurate and reliable when compared with another objective reduction approach based on preserving the dominance relations. To deal with noisy solution sets an eigenvalues based *de-noising* operator is proposed, and it is found to increase the accuracy of the framework. Moreover, to test the robustness of the different objective reduction approaches test problems with and without redundancy are considered, along with two real-world problems.

## 4.1 Proposed Framework for Linear and Nonlinear Objective Reduction

In this section a machine learning based framework is proposed for linear and nonlinear objective reduction algorithms, namely L-PCA and NL-MVU-PCA. The framework is summarised in Framework 1 and aims to find an essential objective set which preserves the correlation-structure of the given non-dominated solution set. If the correlation-structure of the solution set conforms with the correlation-structure of the POF, accurate results are expected. Furthermore, no *a priori* information regarding the nature of the problem is

required and once a non-dominated set is provided the following general steps take place:

1. divide the non-dominated set into objective vectors,
2. find the important directions of variance that are intrinsic in the dataset,
3. identify the conflicting objectives for each direction of variance and build a set with those objectives,
4. identify identically correlated objectives within the previous build set.

---

**Framework 1:** Proposed framework for linear and nonlinear objective reduction

---

**Input:**

$t = 0$  and  $\mathcal{F}_t = \{f_1, f_2, \dots, f_M\}$ .

```

1 begin
2   Obtain a set of non-dominated solutions by running an MOEA corresponding to
    $\mathcal{F}_t$ , for  $N_g$  generations with a population size of  $N$ .
3   Compute a positive semi-definite matrix:  $R$  (Equation 2.11) or  $K$  (Equation 2.15),
   for L-PCA and NL-MVU-PCA, respectively.
4   Compute the eigenvalues and eigenvectors of  $R$  and  $K$  respectively (Section 4.1.2).
5   Perform the eigenvalue analysis (Section 4.1.3) to identify the set of important
   objectives  $\mathcal{F}_e \subseteq \mathcal{F}_t$ .
6   Perform the RCM analysis (Section 4.1.4) to identify the identically correlated
   subsets ( $\mathcal{S}$ ) in  $\mathcal{F}_e$ . If there is no such subset,  $\mathcal{F}_s = \mathcal{F}_e$ .
7   Apply the selection scheme (Section 4.1.5) to identify the most significant
   objective in each  $\mathcal{S}$ , to arrive at  $\mathcal{F}_s$ , such that  $\mathcal{F}_s \subseteq \mathcal{F}_e \subseteq \mathcal{F}_t$ .
8   Compute and store  $\mathcal{E}_t$  (Equation 4.4).
9   if  $\mathcal{F}_s = \mathcal{F}_t$  then
10    | Stop and declare  $\mathcal{F}_t$  as the essential objective set;
11    | Set  $T = t$  and compute the total error  $\mathcal{E}_T$  (Equation 4.5).
12  end
13  else
14    | set  $t = t + 1$ ,  $\mathcal{F}_t = \mathcal{F}_s$ , and go to Step 2.
15  end
16 end

```

---

While the general steps of the framework are based on previous published work in [6] and [7], this new implementation novelty is achieved through the following four goals:

- (i) generality: the framework in [6] and [7] is customised for highly redundant problems, where  $m \ll M$ , while the proposed framework can also be applied to problems with low redundancy, where  $m \approx M$  or  $m = M$ , without loss of generality.
- (ii) de-noising of the input data: the non-dominated set considered for analysis is frequently misrepresentative of the true POF due to the presence of noise in the dataset. The framework handles the possible presence of noise by an eigenvalue based dynamic interpretation of the strength of correlation, which did not exist in [6] and [7].
- (iii) parameter reduction: The number of parameters in the framework have been reduced to a minimum which in turn also reduces the number of possible experiments conducted to evaluate a specific dataset.
- (iv) proposition of an error measure: An error measure is proposed (not in [6] and [7]) so that the variance left unaccounted by the objective reduction results can be accessed and better interpreted.

It needs to be noted that Framework 1 is meant to run iteratively where in each iteration an MOEA runs for  $N_g$  generations with  $\mathcal{F}_t$  as the objective set, and at the end of each iteration an objective reduction takes place. The framework stops iterating when two objective sets deduced as essential by two consecutive iterations, remain the same.

In the following subsections the proposed framework steps are detailed in the following order: (i) construction of a positive semi-definite matrix, namely  $R$  or  $K$  (Section 4.1.1), (ii) eigendecomposition of  $R$  or  $K$  (Section 4.1.2), (iii) eigenvalue analysis (Section 4.1.3), (iv) reduced correlation matrix (RCM) analysis (Section 4.1.4), (v) selection scheme for final reduction based on RCM analysis (Section 4.1.5), and (vi) computation of error (Section 4.1.6).

#### 4.1.1 Construction of a Positive Semi-definite Matrix

A positive semi-definite matrix is built in the first step of the framework from a set of non-dominated solutions corresponding to the initial objective set  $\mathcal{F}_t = \{f_1, \dots, f_M\}$ . In the next

sections, the matrix can be represented by a correlation matrix,  $R$  (Equation 2.11), in case of linear dimensionality reduction, or by a kernel matrix  $K$  (Equation 2.15), in case of nonlinear dimensionality reduction. It should be noted that both matrices have size  $M \times M$ .

### 4.1.2 Eigen-decomposition of the Positive Semi-definite Matrix

After  $R$  or  $K$  are constructed their eigenvalues and eigenvectors are determined by eigen-decomposition. Let the sorted eigenvalues and corresponding eigenvectors be given by  $\lambda_1 \geq \lambda_2 \dots \geq \lambda_M$  and  $V_1, V_2, \dots, V_M$  respectively. At this stage the principal components have been ordered and the following definitions will be utilised in the succeeding sections:

1. Normalise the eigenvalues by  $e_i = \lambda_i / \sum_{i=1}^M \lambda_i$  so that  $\sum_{j=1}^M e_j = 1$ .
2. Consider  $f_{ij}$  as the  $i^{th}$  element of the  $j^{th}$  principal component ( $V_j$ ). Each  $f_i$  can be understood as the contribution towards  $V_j \in \mathbb{R}^M$ .
3. Since the eigenvalues are orthogonal then  $|V_j| = \sum_{i=1}^M f_{ij}^2 = 1 \ \forall j = 1, \dots, M$ .
4. The contribution of each  $f_i$  for all  $V_j$ 's can be given by  $c_i^M = \sum_{j=1}^M e_j f_{ij}^2$ , where  $\sum_{i=1}^M c_i^M = 1$ .

### 4.1.3 Eigenvalue Analysis

The information provided by the eigenvalue analysis is necessary to extract a set of conflicting objectives from the objective set  $\mathcal{F}_t$ . The procedure followed by this analysis is given in the following two steps:

1. The number of principal components ( $N^v$ ) to be considered is determined such that the inequality  $\sum_{i=1}^{N^v} e_i \geq \theta \sum_{i=1}^M e_i$  is respected, where a variance threshold ( $\theta$ ) lies between  $0 \leq \theta \leq 1$ . For this study, the selected  $\theta$  is 0.997<sup>1</sup>.

---

<sup>1</sup>departure from [6] and [7] to make the interpretation scheme more robust. This value is recommended for *any* nature of the distribution of the values in the non-dominated set, though it is chosen in analogy with Gaussian distributions where it accounts for  $\pm 3\sigma$ .

2. In each significant  $V_j$  the objectives considered as important are (i) along each  $V_j$  with the highest absolute and all the opposite-sign elements<sup>2</sup> and (ii) the two best elements, negative or positive, when the  $V_j$  has all the elements with the same sign.

After this step, the set of identified important objectives is denoted by  $\mathcal{F}_e$ , where  $\mathcal{F}_e \subseteq \mathcal{F}_0$ .

#### 4.1.4 Reduced Correlation Matrix (RCM) Analysis

In the following analysis, the possibility of eliminating objectives from  $\mathcal{F}_e$  is measured using: (i) the correlation signs, and (ii) the correlation strengths between objectives. This analysis is performed using the correlation matrix  $R$  and never the kernel matrix  $K$ , that is independent if linear or nonlinear dimensionality reduction is requested. The reason for this recommendation lies in the fact that the correlations existing in  $K$  do not necessarily represent the original relationships between objective functions, which becomes “erroneous” because the isometry of the data is maintained only locally and not globally, during the unfolding process of MVU. Towards it, the process is described in the following two steps:

1. From the correlation matrix  $R$ , build a reduce correlation matrix (RCM). The new RCM matrix is similar to  $R$  except that the columns corresponding to the objectives in  $\mathcal{F}_t \setminus \mathcal{F}_e$ , are not included.<sup>3</sup>
2. Secondly, identify within RCM, the correlated objectives in such a way that two objectives  $f_i, f_j \in \mathcal{F}_e$ , will be considered correlated if the following conditions are satisfied:

$$\begin{aligned} (i) \quad & \text{sign}(R_{ik}) = \text{sign}(R_{jk}) \quad \forall k = 1, 2, \dots, M \text{ and} \\ (ii) \quad & R_{ij} \geq T_{cor}, \text{ where } T_{cor} \text{ is the correlation threshold.} \end{aligned} \tag{4.1}$$

---

<sup>2</sup>departure from [6] and [7], where different principal components were differently interpreted guided by parameters. This makes the current interpretation of principal components parameterless, hence, more robust.

<sup>3</sup>taken from [6] and [7], the rows and columns corresponding to  $\mathcal{F}_t \setminus \mathcal{F}_e$  were excluded from  $R$ . The need for this reduction is captured from the fact that correlation calculated from sets that correspond to: (a) all the objectives  $\mathcal{F}_t \supseteq \mathcal{F}_e$ , and (b) only objectives in  $\mathcal{F}_e$ , are different. Regarding the proposed procedure, the reference objective set is  $\mathcal{F}_t$ , the correlation relationships of any two objectives in  $\mathcal{F}_e$ , should be removed based on the objectives in  $\mathcal{F}_t$ , and should not be restricted to  $\mathcal{F}_e$ . In other words, the rows from  $R$  that correspond to objectives in  $\mathcal{F}_t$  need to be preserved.

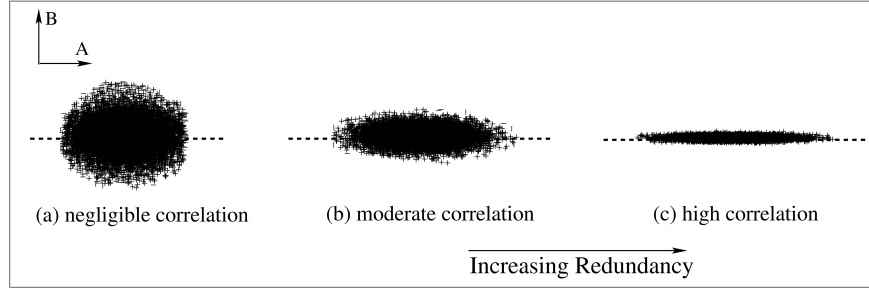


Figure 4.1: A spectrum of possible redundancies in data for two sets of measurements. The best fit line is indicated by a dashed line.

Equation 4.1(i) guarantees that correlation signs from  $f_i$  match with  $f_j$  for all objectives. This comparison involves a set with all the objectives rather than just individual objective values, which brings more meaning to identical correlated objectives since the correlations between objectives are interpreted as a set-based property. In this way, for each  $f_i \in \mathcal{F}_e$  let  $\hat{\mathcal{S}}_i$  denote a *potentially identically* correlated subset which includes all  $f_j \in \mathcal{F}_e$  that satisfy the first condition. The second condition in Equation 4.1(ii) requires the strength of correlation between  $f_i$  and  $f_j$  to be superior to some threshold, denoted by  $T_{cor}$ . Based on this, let those  $f_j, j \in \hat{\mathcal{S}}_i$  that in addition satisfy the second condition constitute an *identically* correlated subset  $\mathcal{S}_i$ .

The remaining discussion in this section describes  $T_{cor}$ . In that, the strength of correlation can be interpreted in many different ways. For instance, the Cohen scale that considers 0.1 to 0.3 as *weak*, 0.3 to 0.5 as *moderate* and 0.5 to 1.0 as *strong*. However, the method here proposed determines  $T_{cor}$  based on the information provided by the problem structure, using the eigenvalues. In order to exemplify how the eigenvalues may produce different results, Figure 4.1 displays a spectrum of different redundancies in the dataset corresponding to two sets of measurements, namely A and B. In the first case, Figure 4.1(a), the sets A and B are totally uncorrelated, and considered low redundancy, and in contrast, Figure 4.1(c), set A is highly correlated with B, and considered high redundancy. Figure 4.1(b) is an example of moderate redundancy compared to the other two examples. Additionally, the difference in the magnitude of the eigenvalues increases as the degree of correlation increases from case

(a) to case (c). For case (a), the eigenvalues are comparable in terms of magnitude, they are unequal for case (b), and extremely different for case (c). This observation regarding the eigenvalue magnitude is also valid for higher dimensional space. In order for this scheme to account for no or low redundancy, medium redundancy and high redundancy cases,  $T_{cor}$  is proposed in Equation 4.2 so that: (i) the proportion of the first eigenvalue with the sum of all eigenvalues, i.e.,  $e_1$ , and (ii) the number of eigenvalues (say,  $M_{2\sigma}$ ) required to account for 95.4% of the variance<sup>4</sup> can be distinguished.

$$T_{cor} = 1.0 - e_1(1.0 - M_{2\sigma}/M) \quad (4.2)$$

For problems with high redundancy,  $e_1$  is also very high, and  $M_{2\sigma}$  will be small compared to  $M$ . On the other hand  $T_{cor}$  will have a small value. For problems with low redundancy,  $e_1$  will be low, and  $M_{2\sigma}$  will be comparable to  $M$ , while  $T_{cor}$  will have a very high value. The proposed formula attributes to  $T_{cor}$  a low value for problems with high redundancy and a high value for problems with low redundancy.

#### 4.1.5 Selection Scheme for Final Reduction based on RCM Analysis

Previously, a subset or subsets of identically correlated objectives in  $\mathcal{F}_e$  were identified by the RCM analysis. On this final reduction, the aim is to identify from each subset of identically correlated objectives the most significant objectives and eliminate the remaining ones.

First, let us consider a particular identically correlated subset  $\mathcal{S}_k$ , where  $k \in \{1, 2, \dots, M\}$  and each objective belonging to this set is  $f_i, i \in \mathcal{S}_k$ . It is already known from Section 4.1.3 that  $V_j$  and  $e_j$  where  $j = 1 \dots N_v$ , represent the principal components and the corresponding eigenvalues that account for the threshold variance  $\theta$ . Considering this, the selection score

---

<sup>4</sup>selected in agreement with the three-sigma rule, where for a Gaussian distribution, 95.4% of the values lie within two standard deviations of the mean, i.e.,  $\pm 2\sigma$ .



attributed to each  $f_i, i \in \mathcal{S}_k$  is given by

$$sc_i = \sum_{j=1}^{N_v} e_j |f_{ij}|, \quad (4.3)$$

where  $f_{ij}$  represents the contribution of  $f_i$  along  $V_j$ .

The value of  $sc_i$  accounts for the given contribution by each  $f_i$  along the considered principal components. For the highest value of  $sc_i$  in the set  $\{sc_i, i \in \mathcal{S}_k\}$  let the corresponding objective  $f_i$  denote the most significant objective in  $\mathcal{S}_k$ . Then, retain the most significant objective belonging to each identically correlated subset and eliminate the other objectives. When all significant objectives are identified the set created from this reduction is denoted as  $\mathcal{F}_s$ . This final set is called the *essential* objective set after one iteration of the proposed framework.

#### 4.1.6 Computation of Error

A measurement for calculating the error incurred, after an objective reduction takes place is proposed. The error measures the variance that is left unaccounted when objectives in  $\mathcal{F}_{redn} = \mathcal{F}_0 \setminus \mathcal{F}_s$  are discarded. The error computation formulation is given by

$$\left. \begin{aligned} \mathcal{E}_t &= \sum_{i \in \mathcal{F}_{redn}} c_i^M \left( 1.0 - \max_{j \in \mathcal{F}_s} \{\delta_{ij} \cdot R_{ij}\} \right) \\ \text{where} \\ c_i^M &= \sum_{k=1}^M e_k f_{ik}^2 \\ \delta_{ij} &= \begin{cases} 1, & \text{if } f_i \text{ and } f_j \text{ are identically correlated} \\ 0, & \text{otherwise} \end{cases} \\ R_{ij} &= \text{strength of correlation between } f_i \text{ and } f_j \end{aligned} \right\}. \quad (4.4)$$

The rationale behind Equation 4.4 is explained in the following two points:

1. The variance of an objective  $f_i \in \mathcal{F}_{redn}$  that is not identically correlated with any other

objective  $f_j \in \mathcal{F}_s$ , i.e.  $\delta_{ij} = 0$ , is  $c_i^M$ .

2. When an objective  $f_i \in \mathcal{F}_{redn}$  is identically correlated with any objective  $f_j \in \mathcal{F}_s$  part of the variance is already accounted by objective  $f_j$ . Hence, the variance left unaccounted for the removal of objective  $f_i$  is reduced by a fraction of  $R_{ij}$ .

The framework runs iteratively for  $T$  iterations, where in each iteration the error computed is given by Equation 4.4. However, the total error that is incurred when successive reductions take place in each iteration is given by Equation 4.5. The rationale to compute the total error is as follows: (i) in the first iteration the error  $\mathcal{E}_0$  corresponds to the removal of  $\mathcal{F}_{redn}$  from  $\mathcal{F}_0$ , and (ii) during the second iteration the variance that is left unaccounted for is  $(1 - \mathcal{E}_{t-1})$ , hence the error reported in  $\mathcal{E}_t$  is only a fraction of the that variance. The generalisation for  $T$  iteration of this rational is given by

$$\mathcal{E}_T = \mathcal{E}_0 + \sum_{t=1}^T \mathcal{E}_t (1 - \mathcal{E}_{t-1}). \quad (4.5)$$

## 4.2 Generality and Efficiency of the Framework

In this section a summary of the enhancements proposed over [6] and [7] is discussed around the efficiency and generality of the framework.

The general steps followed by the framework are common with those reported in [6] and [7], however their implementation in the current framework differs so that four main goals are achieved as summarised in Table 4.1 and detailed below:

1. adaptation: in eigenvalue analysis (Section 4.1.3) the replacement of a threshold cut (TC) by  $\theta = 0.997$  as recommended in Section 4.1.3, and a more robust principal components interpretation (culminating in a more sensitive  $\mathcal{F}_e$ ) allows a better handling of redundancy.
2. correction: in [6] and [7] both the rows and columns were removed from the reduced

Table 4.1: Classification of proposed enhancements, over [6, 7]

Sections	Status in [6, 7]	Adaptation	Correction	Addition	Parameter reduction
4.1.1) $q$	A parameter				✓
4.1.3) $\theta$	Suitable for $m \ll M$	✓			
$\mathcal{F}_e$	Parameter dependent, suitable for $m \ll M$	✓			✓
4.1.4) RCM	Erroneous <sup>a</sup>		✓		
Eq. 4.1(i)			✓		
Eq. 4.1(ii)	Absent			✓	
4.1.5) Eq. 4.3	Absent <sup>b</sup>			✓	
4.1.6) Eq. 4.4	Absent			✓	

<sup>a</sup> The manner in which the reduced correlation matrix (RCM) was constructed, did not allow for correlation to be treated as a set based property.

<sup>b</sup> It was pursued on an ad hoc basis.

correlation matrix (RCM) and now only the columns are removed. This acknowledges that correlations should be interpreted based on the entire set of objectives by allowing it to be treated as a set based property.

- addition: the additions are (i) the correlation threshold ( $T_{cor}$  in Equation 4.2) which serves to have a de-noising effect, (ii) the objective contribution score ( $sc_i$  in Equation 4.3) that helps to select objectives during the RCM analysis, and (iii) the error measure ( $\mathcal{E}_t$  in Equation 4.4) for objective reduction justification.
- parameter reduction: it is recommended to set  $q = M - 1$  (Section 4.1.1) for all purposes since a sufficient degree of freedom is given to the unfolding process, while preserving the local isometry. The threshold cut (TC) in [6] and [7] used during eigenvalue analysis (Section 4.1.3) has been removed.

Table 4.2 shows how the framework is able to handle problems with different redundancy and also how it can be customised for increasing its efficiency. For highly redundant problems, where  $m \ll M$ , the selection of: (i) a low  $\theta$ , (ii) less number of objectives picked by each principal component and (iii) a lower  $T_{cor}$  can be used. For problems with low redundancy,

Table 4.2: Handling of problem redundancy, over [6, 7]

Sections	Status in [6, 7]	Conflicting settings for problems with:	
		$m \ll M$	$m \approx M$
4.1.3) $\theta$	Suitable for $m \ll M$	Low	High
$\mathcal{F}_e$	Parameter dependent, suitable for $m \ll M$	Few objectives per principal component so that: $ \mathcal{F}_e  \ll  \mathcal{F}_0 $	More objectives per principal component so that: $ \mathcal{F}_e  \approx  \mathcal{F}_0 $
4.1.4) Eq. 4.1(ii)	Absent	Low	High

where  $m \approx M$  or  $m = M$ , the eigenvalues analysis reduction can be skipped and the RCM analysis can be performed with a higher value of  $T_{cor}$ . When the framework is customised for specific problems it can incur in a loss of generality and the performance may be affected if the problem redundancy is initially unknown. For instance, in case the framework is customised for problems where  $m \ll M$  it may perform poorly in problems with  $m \approx M$  or  $m = M$ , and vice-versa.

The proposed framework does not require any prior information about the provided dataset nature. It is therefore very important to keep it as general as possible so that the final results can be trusted. It adopts a high value of  $\theta$  and a uniform approach of composing  $\mathcal{F}_e$ , but dynamically assigns  $T_{cor}$  based on the problem information revealed by the eigenvalues. Figure 4.2 summarises the above discussions.

### 4.3 Test Problems

In this section the test problems are differentiated as redundant and non-redundant, depending on the redundancy manifested between their objectives. Also, their characteristics and features relevant for this chapter are described in the following subsections.

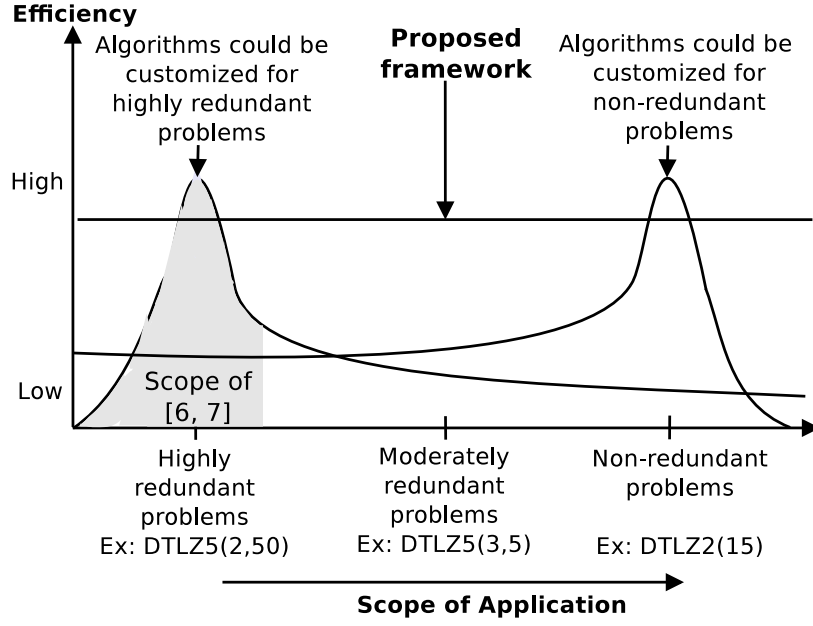


Figure 4.2: Highlighting the scope and efficiency of the proposed framework.

### 4.3.1 Redundant Problems

The redundant test problems are comprised of different versions of DTLZ5( $I, M$ ) [6] (variant of DTLZ5 problem [27]) and WFG3( $M$ ) [29] scalable problems.

The relevant characteristics for DTLZ5( $I, M$ ) are:

1. for an  $M$ -objective problem the dimensions of the POF is determined by  $I$ , where  $I < M$ .
2. the identically correlated objectives (non-conflict) are the first  $M - I + 1$  while the remaining are in conflict with every other objective in the problem. The correlation-structure captured is given by

$$\text{sign}(R_{ij}) = \begin{cases} + & i, j = 1 : M - I + 1 \\ + \quad j = i \\ - \quad j \neq i \end{cases} \quad i, j = M - I + 2 : M \quad (4.6)$$

where  $R_{ij} = 1.0 \forall i, j = 1 : M - I + 1$ .

3. an essential objective set is given by

$$\mathcal{F}_{\mathcal{T}} = \{f_k, f_{M-I+2}, \dots, f_M\}, \quad (4.7)$$

where  $k \in \{1, \dots, M - I + 1\}$  in general and  $k = M - I + 1$  for L-PCA and NL-MVU-PCA in particular. The objective  $f_k$  for L-PCA and NL-MVU-PCA is selected based on the largest variance between the possible  $ks$ , hence  $k = M - I + 1$ .

The relevant characteristics for WFG3( $M$ ) are:

1. the problem POF degenerates into a linear hypercube leading to  $\sum_{m=1}^M f_m = 1$ . This causes the first  $M - 1$  objectives to be perfectly correlated and conflict exists only between the last objective with all the others. The correlation-structure is given by

$$\text{sign}(R_{ij}) = \left\{ \begin{array}{ll} + & i, j = 1 : M - 1 \\ - & j = 1 : M - 1 \\ + & j = M \end{array} \right\} \quad i = M \quad (4.8)$$

where  $R_{ij} = 1.0 \ \forall \ i, j = 1 : M - 1$ .

2. an essential objective set is given by

$$\mathcal{F}_{\mathcal{T}} = \{f_k, f_M\}, \quad (4.9)$$

where  $k \in \{1, \dots, M - 1\}$  in general and  $k = M - 1$  for L-PCA and NL-MVU-PCA in particular. The objective  $f_k$  is selected among the  $ks$  with higher variance for NL-MVU-PCA and L-PCA, leading to  $k = M - 1$ .

### 4.3.2 Non-Redundant Problems

The non-redundant problems are comprised of four versions of the DTLZ scalable test problems [27], namely DTLZ1, DTLZ2, DTLZ3 and DTLZ4. The relevant problems' character-

istics are:

1. for an  $M$ -objective problem, when the solution set is considered optimal, the dimensions of the POF is defined by  $M$ .
2. for these problems all the objectives are in conflict which leads into the correlation-structure given by

$$\text{sign}(R_{ij}) = \begin{cases} + & i = j \\ - & i \neq j \end{cases} \quad i, j = 1 : M. \quad (4.10)$$

3. since all objectives are considered to be important the essential objective set is given by

$$\mathcal{F}_{\mathcal{T}} = \{f_1, \dots, f_M\}. \quad (4.11)$$

## 4.4 Quality Indicators

In this section, two quality indicators (also performance measures) for comparing non-dominated solution sets are described. The quality indicators are considered essential to evaluate the performance of different MOEAs. One quality indicator is used to measure convergence, i.e. how distant the solutions are from the POF. The second one measures diversity, i.e., the spread out across the POF. Note that, the application of these quality indicators is only possible because the POF of the considered test problems is known.

Before describing the quality indicators, note that the test problems parameters can be categorised by either *distance*, *position*, or *mixed* parameters, as suggested in [29]. The distance parameters affect the convergence of the solutions while a position parameter affects the position of the solution along the same Pareto-front. A mixed parameter can affect the convergence and also the position of the solution. For the considered test problems the parameters can be either *distance* or *position*, and let their number be denoted by  $\kappa$  and  $\rho$ ,

Table 4.3:  $g$  function for DTLZ and DTLZ5( $I, M$ ) test problems

Name	$g$ function	$\kappa$	Ideal $g$ value
DTLZ1	$100 \left[  X  + \sum_{x_i \in X} (x_i - 0.5)^2 - \cos(20\pi(x_i - 0.5)) \right]$	5	$g = 0$
DTLZ2	$\sum_{x_i \in X} (x_i - 0.5)^2$	10	$g = 0$
DTLZ3	same as DTLZ1	10	$g = 0$
DTLZ4	same as DTLZ2	10	$g = 0$
DTLZ5( $I, M$ )	same as DTLZ2	10	$g = 0$

respectively.

To evaluate the convergence of the non-dominated solution set, in here the  $g$  function is used [27]. This function corresponds to the minimum value of the Pareto-optimal surface obtained for DTLZ [27] and DTLZ5( $I, M$ ) [6] test problems. For each problem, the formulation of  $g$  is given in Table 4.3, where  $M$  represents the number of objectives and  $X$  the *distance* parameters. Additionally, it is also mentioned the number of *distance* parameter ( $\kappa$ ) and the the ideal  $g$  value per problem. Notably, the ideal  $g$  value is obtained only when the non-dominated solution set is on the POF.

For diversity, the normalised maximum spread indicator ( $I_s$ ) [60] is used and the formulation is given by

$$I_s = \frac{\mathcal{D}_A}{\mathcal{D}_T} = \frac{\left[ \sum_{i=1}^M \left( \max_{f \in Z_A} f_i - \min_{f \in Z_A} f_i \right)^2 \right]^{\frac{1}{2}}}{\left[ \sum_{i=1}^M \left( \max_{f \in Z_T} f_i - \min_{f \in Z_T} f_i \right)^2 \right]^{\frac{1}{2}}}, \quad (4.12)$$

where  $\mathcal{D}_A$  is the obtained dispersal of solutions and  $\mathcal{D}_T$  is the dispersal of solution on the true POF. This quality indicator helps to assess the dispersal of solutions in  $Z_A$  (set of obtained solutions) against the ideal in  $Z_T$  (set of solutions on the true POF), as follows:

- (i) for  $I_s > 1$ , the measure indicates that solutions are dispersed in regions which are not part of the global trade-off surface;



Table 4.4: Derivations for  $\mathcal{D}_{\mathcal{T}}$  (dispersal of solutions on the true POF)

$\mathcal{D}_{\mathcal{T}}$  is used to compute  $I_s$  as given by Equation 4.12. For brevity, the test problems names denoted by  $Dk$  and  $D5(I, M)$  correspond to DTLZ $k$  and DTLZ5( $I, M$ ), respectively

Name	$f^{max}$	$f^{min}$	$\mathcal{D}_{\mathcal{T}}^2 = \sum_{i=0}^M (f_i^{max} - f_i^{min})^2$
D1	$f_i = \frac{1}{2}, \forall i = 1, \dots, M$	$f_i = 0, \forall i = 1, \dots, M$	$\sum_{i=1}^M \left(\frac{1}{2}\right)^2 = 0.25M$
D2			
D3	$f_i = 1, \forall i = 1, \dots, M$	$f_i = 0, \forall i = 1, \dots, M$	$\sum_{i=1}^M 1 = M$
D4			
D5( $I, M$ )	$f_1 = \left(\frac{1}{\sqrt{2}}\right)^{M-I}$ $f_i = \left(\frac{1}{\sqrt{2}}\right)^{M-I+2-i}, \forall i = 2, \dots, M - I + 1$ $f_i = 1, \forall i = M - I + 2, \dots, M$	$f_i = 0, \forall i = 1, \dots, M$	Note*
* $\left(\frac{1}{2}\right)^{M-I} + \sum_{i=2}^{M-I+1} \left(\frac{1}{2}\right)^{M-I+2-i} + I - 1$			

- (ii) for situations where  $Z_A$  converged to sub-regions which are possible global optimal, the indication is given by  $I_s < 1$ .

In this way,  $I_s$  directly indicates if there is too much or too little spread of solutions, away or over the POF. The derivations of  $\mathcal{D}_{\mathcal{T}}$  are given in Table 4.4 for DTLZ and DTLZ5( $I, M$ ) test problems.

## 4.5 Experimental Settings

To evaluate the performance of the proposed objective reduction algorithms the following non-dominated solution sets are used:

1. solutions sampled on the true POF, denoted by  $\mathcal{N}_{\mathcal{P}}$ , represent *unnoised* signal. To generate these solutions the sampling scheme in [44] has been used. The number of solutions is given by  $N = C_{m-1}^{H+m-1}$ , where  $H$  is a user defined parameter and  $m$  represents the dimension of the POF. In here, for  $m \leq 10$  :  $H = 8$  and for  $m > 10$  :  $H = 4$  is used. The solutions are generated only in one run since the method is deterministic.
2. solutions generated by MOEAs that represent a mixture between unnoised and noised

signal. These solutions have been obtained by: (i) NSGA-II [31], denoted by  $\mathcal{N}_{\mathcal{NS}}$ , or (ii)  $\epsilon$ -MOEA [34], denoted by  $\mathcal{N}_{\epsilon}$ . For both cases the number of generations is 2000 and the population size is 200. The probability of crossover and mutation is 0.9 and 0.1, respectively, and the distribution index for crossover and mutation is 5 and 20, respectively. Specifically for  $\epsilon$ -MOEA,  $\epsilon = 0.3$  is used. The same parameters have been selected by [6] and the only exception is the probability of mutation where the respective authors have selected  $1/n$ , ( $n$  is the number parameters in the test problem). In here, it has been decided to fix the probability of mutation for all test problems since it has been found experimentally to contribute to the dispersion of the population even when  $n$  is high. Moreover, the solutions have been generated using 20 random equally spaced seeds.

3. solutions randomly sampled on the search space, denoted by  $\mathcal{N}_{\mathcal{R}}$ , represent *noised* signal. These solutions correspond to the random initialised population generated by NSGA-II with a size of 200.

Moreover, the number of *distance* parameters for DTLZ and DTLZ5( $I, M$ ) test problems have already been mentioned in Table 4.4, while for WFG3 the number is set to 20. The number of *position* parameters is set to  $\rho = M - 1$  for all test problems.

## 4.6 Feature Selection by the Proposed Algorithms

During eigenvalues analysis the proposed algorithms determine objectives to be conflicting or non-conflicting based on their signs along the principal components. Two objectives are in conflict when they have opposite signs, otherwise they are non-conflicting. This procedure is demonstrated in practice for DTLZ5(2, 3), using  $\mathcal{N}_{\mathcal{NS}}$ , where the principal components ( $V$ ) together with respective normalised eigenvalues ( $e$ ) are shown in Table 4.5 and plotted in Figure 4.3a.

Notably, the first principal component ( $V_1$ ) accounts for 96.1% of the variance while the

Table 4.5: Correlation matrix  $R$ : Use of PCA for feature selection demonstrated on DTLZ5(2, 3).

	$e_1 = 0.961$	$e_2 = 0.038$	$e_3 \approx 0$
	$V_1$	$V_2$	$V_3$
$f_1$	0.583	0.400	0.707
$f_2$	0.583	0.400	-0.707
$f_3$	-0.565	0.825	0.000

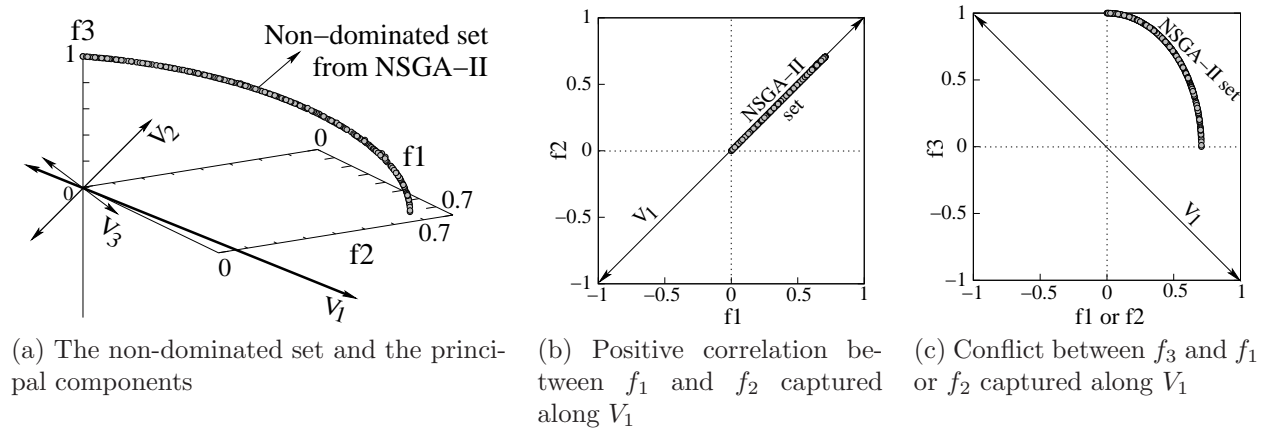


Figure 4.3: Non-dominated set against principal components: Use of PCA for feature selection demonstrated on DTLZ5(2, 3).

contributions of  $f_1$  and  $f_2$  are positive (their sign is positive), and the contribution of  $f_3$  is negative (negative sign). The principal components interpretation is validated by: (i) Figure 4.3b since objectives  $f_1$  and  $f_2$  are both increasing or decreasing together along  $V_1$  and (ii) Figure 4.3c depicts an increase/decrease in  $f_3$  calls for a decrease/increase in both  $f_1$  and  $f_2$ . In this problem it is obvious that  $f_1$  and  $f_2$  are non-conflicting objectives, hence, either  $f_1$  or  $f_2$  is a redundant objective and that fact is captured by the proposed feature selection approach.

The previous analysis was conducted using DTLZ5(2, 3) test problem where the objectives redundancy could be visually captured by Figure 4.3a. Also the non-dominated set conforms totally with the POF and hence there is no visible noise in the dataset. In reality the presence of noise in non-dominated sets obtained by MOEAs is common, especially when the number of objectives increases. It is also very difficult to visually identify the objectives redundancy when the number of objectives is greater than three. This corroborates the need for an analytical tool that identifies the conflicts between objectives which is provided by proposed algorithms, as demonstrated for DTLZ5(3, 5) in the subsequent sections.

## 4.7 Working of the Proposed Algorithms when Applied to Data Sets Without Noise

Here a discussion is conducted on how a feature selection based approach operates for the considered algorithms during objective reduction. The algorithms are then demonstrated on DTLZ5(3, 5) when applied to  $\mathcal{N}_{\mathcal{P}}$ . Due to moderate redundancy the test problem DTLZ5(3, 5) is selected for the proposed framework evaluation.

### 4.7.1 DTLZ5(3,5): L-PCA and NL-MVU-PCA based on $\mathcal{N}_{\mathcal{P}}$

In this section the linear and nonlinear algorithms, L-PCA and NL-MVU-PCA, are applied to  $\mathcal{N}_{\mathcal{P}}$ . The correlation matrix  $R$  and the kernel matrix  $K$  are shown in Tables 4.6a and 4.6b,

Table 4.6: DTLZ5(3,5): The  $R$  and  $K$  matrix with their corresponding eigenvalues and eigenvectors, for  $\mathcal{N}_{\mathcal{P}}$  (one run)

(a) Correlation matrix  $R$

	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
$f_1$	1.0000	1.0000	1.0000	-0.4749	-0.4340
$f_2$	1.0000	1.0000	1.0000	-0.4749	-0.4340
$f_3$	1.0000	1.0000	1.0000	-0.4749	-0.4340
$f_4$	-0.4749	-0.4749	-0.4749	1.0000	-0.4340
$f_5$	-0.4340	-0.4340	-0.4340	-0.4340	1.0000

(b) Kernel matrix  $K$

	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
$f_1$	2.5862	2.5862	4.0156	-4.7231	-4.4650
$f_2$	2.5862	2.5862	4.0156	-4.7231	-4.4650
$f_3$	4.0156	4.0156	6.3176	-7.4018	-6.9470
$f_4$	-4.7231	-4.7231	-7.4018	23.0615	-6.2134
$f_5$	-4.4650	-4.4650	-6.9470	-6.2134	22.0906

(c) Eigenvalues and respective eigenvectors of  $R$

	$e_1$	$e_2$	$e_3$	$e_4$	$e_5$
	0.6867	0.2866	0.0266	0.0000	0.0000
	$V_1$	$V_2$	$V_3$	$V_4$	$V_5$
$f_1$	0.538	-0.009	0.209	0.405	0.708
$f_2$	0.538	-0.009	0.209	0.410	-0.705
$f_3$	0.538	-0.009	0.209	-0.816	-0.002
$f_4$	-0.272	-0.691	0.668	0.000	0.000
$f_5$	-0.239	0.722	0.649	0.000	0.000

(d) Eigenvalues and respective eigenvectors of  $K$

	$e_1$	$e_2$	$e_3$	$e_4$	$e_5$
	0.5138	0.4855	0.0006	0.0000	0.0000
	$V_1$	$V_2$	$V_3$	$V_4$	$V_5$
$f_1$	0.134	0.273	-0.455	-0.714	0.435
$f_2$	0.134	0.273	-0.455	-0.087	-0.832
$f_3$	0.211	0.426	0.757	-0.400	-0.198
$f_4$	-0.874	-0.173	0.077	-0.400	-0.198
$f_5$	0.394	-0.799	0.075	-0.400	-0.198

respectively, while by eigendecomposition the corresponding eigenvalues and eigenvectors of  $R$  and  $K$  are shown in Tables 4.6c and 4.6d, respectively. Using the provided information each algorithm analysis is as follows.

### DTLZ5(3,5): L-PCA based on $\mathcal{N}_{\mathcal{P}}$

Using the eigenvalues and eigenvectors provided by the correlation matrix  $R$ , the remaining L-PCA steps are summarised in Table 4.7. The analysis that follows is based on  $\mathcal{N}_{\mathcal{P}}$  for DTLZ5(3,5) using the L-PCA algorithm and the steps are:

1. Vector  $V_1$  in Table 4.6c is the first principal component where most problem variance is accounted (at least 68.67% since  $e_1 = 0.6867$ ), whose components define a five-dimensional ray in the objective space that represents the contribution of each objective. Considering the interpretation of the principal components based on their signs (given in Section 4.6), for the given example, objectives  $f_1$ ,  $f_2$  and  $f_3$  (positive sign) can be interpreted as being in conflict with  $f_4$  and  $f_5$  (negative sign) due to their signs in Table 4.6. In Table 4.7a the eigenvalues analysis selected the top three prin-

Table 4.7: DTLZ5(3, 5): Iteration1 of L-PCA with  $\mathcal{N}_{\mathcal{P}}$  (one run)

(a) Eigenvalue analysis							
PCA	Variance	Cumulative	Objectives Selected				
$(N^v)$	(%)	(%)	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
1	68.67	68.67			$f_3$	$f_4$	$f_5$
2	28.66	97.33	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
3	02.66	99.99				$f_4$	$f_5$

(b) RCM analysis			
1) Correlation over whole set	Eq 4.1(i)	$\hat{\mathcal{S}}_1 = \hat{\mathcal{S}}_2 = \hat{\mathcal{S}}_3 = \{f_1, f_2, f_3\}$	
2) Correlation threshold: $T_{cor}$	Eq 4.2	$1.0 - 0.6867 \times (1.0 - 2/5) = 0.5879$	
3) Correlation meeting $T_{cor}$	Eq 4.1(ii)	$\mathcal{S}_1 = \mathcal{S}_2 = \mathcal{S}_3 = \{f_1, f_2, f_3\}$	

(c) Selection scheme				
	$e_1 = 0.6867$	$e_2 = 0.2866$	$e_3 = 0.0266$	
	$V_1$	$V_2$	$V_3$	$sc_i$
$f_1$	0.538	-0.009	0.209	0.3778
$f_2$	0.538	-0.009	0.209	0.3778
$f_3$	0.538	-0.009	0.209	0.3778

principal components since their total variance is higher than  $\theta = 0.997$ . From selected principal components all the objectives are picked leading to  $\mathcal{F}_e = \{f_1, f_2, f_3, f_4, f_5\}$ .

- The interpretation of the correlation signs in Table 4.6a has led the RCM analysis (Table 4.7b) into three potential correlated sets, namely  $\hat{\mathcal{S}}_1 = \hat{\mathcal{S}}_2 = \hat{\mathcal{S}}_3 = \{f_1, f_2, f_3\}$ , since the correlation signs of  $f_1$  are equal to those in  $f_2$  and as well  $f_3$  columns. After obtaining  $T_{cor} = 0.5879$  the potential correlated sets become indeed correlated since their correlations  $R_{12} = R_{13} = R_{23} = 1.0$ . The RCM analysis last step confirms the previous potential correlated as indeed correlated since their correlations  $R_{12} = R_{13} = R_{23} = 1.0$  are higher than the obtained  $T_{cor} = 0.5879$ , hence  $\mathcal{S}_1 = \mathcal{S}_2 = \mathcal{S}_3 = \{f_1, f_2, f_3\}$ .
- The selection scheme in Table 4.7c considers that any objective from the correlated sets can be picked, since their objective selection scores are identical.

Moreover, L-PCA determines  $\mathcal{F}_s = \{f_k, f_4, f_5\}$  where  $k$  can be any objective from the set  $\{f_1, f_2, f_3\}$ . In case  $k = 3$  leading to  $\mathcal{F}_s = \{f_3, f_4, f_5\}$  the variance that is left unaccounted for (corresponding to the error measure proposed in Equation 5.3) by discarding objectives  $\{f_1, f_2\}$  is  $\mathcal{E}_0 = c_1^M(1.0 - R_{13}) + c_2^M(1.0 - R_{23}) = 0.0$ . The second iteration could not reduce the objective set further, therefore it is not shown here.

### DTLZ5(3,5): NL-MVU-PCA based on $\mathcal{N}_{\mathcal{P}}$

The analysis that follows is based on  $\mathcal{N}_{\mathcal{P}}$  for DTLZ5(3, 5) using the NL-MVU-PCA algorithm, as represented in Table 4.8. The steps are:

1. The eigenvalues and eigenvectors are obtained from the kernel matrix, in Table 4.6b, and reported in Table 4.6d. The eigenvalues analysis in Table 4.8a considers the top two principal components as essential and the objectives selected by them leads to  $\mathcal{F}_e = \{f_1, f_2, f_3, f_4, f_5\}$ , which means that no objective reduction takes place at this stage.
2. It should be noted that, for nonlinear dimensionality reduction procedure, the matrix  $K$  is only used for finding the principal components and the RCM analysis is based on the correlation matrix  $R$ . The RCM analysis in Table 4.8b considers the possible correlated sets as  $\hat{\mathcal{S}}_1 = \hat{\mathcal{S}}_2 = \hat{\mathcal{S}}_3 = \{f_1, f_2, f_3\}$  since their correlations signs in  $R$  (Table 4.6a) along their columns are identical. The objectives in the possible correlated sets are considered truly correlated since their correlation values  $R_{12} = R_{13} = R_{23} = 1.0$  are higher than the obtained threshold  $T_{cor} = 0.7086$ , hence,  $\mathcal{S}_1 = \mathcal{S}_2 = \mathcal{S}_3 = \{f_1, f_2, f_3\}$ .
3. The selection scheme analysis in Table 4.8c picks objective  $f_3$  as the most conflicting from the correlated set, leading to the essential objective set  $\mathcal{F}_s = \{f_3, f_4, f_5\}$ .
4. The error obtained by the removed objective set  $\{f_1, f_2\}$  is  $\mathcal{E}_0 = c_1^M(1.0 - R_{13}) + c_2^M(1.0 - R_{23}) = 0.0$ . Since during the second iteration, no further reduction was possible, it is not shown.

Table 4.8: DTLZ5(3, 5): Iteration1 of NL-MVU-PCA with  $\mathcal{N}_{\mathcal{P}}$  (one run)

(a) Eigenvalue analysis							
PCA	Variance	Cumulative	Objectives Selected				
$(N^v)$	(%)	(%)	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
1	51.38	51.38	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
2	48.55	99.93	$f_1$	$f_2$	$f_3$		$f_5$

(b) RCM analysis			
1) Correlation over whole set	Eq 4.1(i)	$\hat{\mathcal{S}}_1 = \hat{\mathcal{S}}_2 = \hat{\mathcal{S}}_3 = \{f_1, f_2, f_3\}$	
2) Correlation threshold: $T_{cor}$	Eq 4.2	$1.0 - 0.5138 \times (1.0 - 2/5) = 0.7086$	
3) Correlation meeting $T_{cor}$	Eq 4.1(ii)	$\mathcal{S}_1 = \mathcal{S}_2 = \mathcal{S}_3 = \{f_1, f_2, f_3\}$	

(c) Selection scheme			
	$e_1 = 0.5138$	$e_2 = 0.4855$	
	$V_1$	$V_2$	$sc_i$
$f_1$	0.134	0.273	0.2017
$f_2$	0.134	0.273	0.2017
$f_3$	0.211	0.426	0.3154

Following the described steps NL-MVU-PCA identifies the essential objective set  $\mathcal{F}_t = \{f_3, f_4, f_5\}$  and  $\mathcal{F}_{redn} = \{f_1, f_2\}$ , which incurs in error  $\mathcal{E}_t = 0$ .

#### 4.7.2 Key Inferences from the Analysis based on $\mathcal{N}_{\mathcal{P}}$

The results shown above have highlighted the following DTLZ5(3, 5) problem characteristics:

1. The objectives  $\{f_1, f_2, f_2\}$  are positively correlated with a strength of correlation 1.0 as reported in Table 4.6a.
2. The variance of objective  $f_3$  captured by L-PCA is equal to objectives  $f_1$  and  $f_2$ , since the selection scores are  $sc_3 = sc_2 = sc_1$ , as shown in Table 4.7c. On the other hand, NL-MVU-PCA has captured higher variance by  $f_3$  when compared with objectives  $f_1$  and  $f_2$ , since the selection scores are  $sc_3 > sc_2 = sc_1$ , as shown in Table 4.8c. The latter is corroborate by the parallel coordinate plot<sup>5</sup> shown in Figure 4.4 where it is

---

<sup>5</sup>Parallel coordinate plots is a high-dimensional visualisation technique presented in [158] that is widely used to analyse multivariate data.



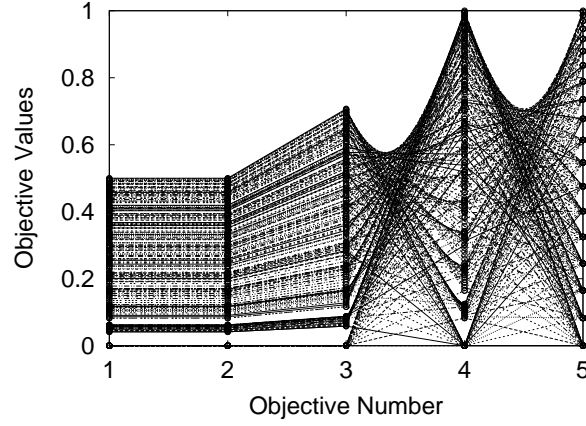


Figure 4.4: DTLZ5(3, 5): Parallel coordinate plot for  $\mathcal{N}_{\mathcal{P}}$ .

possible to visualise that the variance of objective  $f_3$  is higher than objectives  $f_1$  and  $f_2$ .

Since the PCA method is based on the premise that higher variance in the data represents the signal, then, in a problem like DTLZ5(3, 5), objective  $f_3$  should be preferred to  $f_1$  and  $f_2$ . In the proposed algorithms, L-PCA and NL-MVU-PCA, an essential objective is selected from each identically correlated objective set using the selection score (Equation 4.3). For the problem in consideration the selection scores captured should be  $sc_3 > sc_2 = sc_1$ , for  $f_3$  to be selected. As a matter of fact L-PCA fails to capture the superiority of  $f_3$  since all selection scores are equal (Table 4.7c). On the contrary, NL-MVU-PCA acknowledged  $sc_3$  as superior and both  $sc_2$  and  $sc_1$  as equal (Table 4.8c). The reasons beyond PCA failure are attributed to the captured principal components as shown in Table 4.6c. In that respect the variance contribution of  $f_1$ ,  $f_2$  and  $f_3$  are erroneously equally considered. For this example the deficit in L-PCA accuracy in objective reduction is attributed to the inaccurate principal components that failed to capture the problem variance correctly.

Table 4.9: DTLZ5(3,5): The  $R$  and  $K$  matrix with their corresponding eigenvalues and eigenvectors, for  $\mathcal{N}_{\mathcal{NS}}$  (one run)

(a) Correlation matrix $R$						(b) Kernel matrix $K$					
	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$		$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
$f_1$	1.0000	0.5668	0.6880	-0.2852	-0.4257	$f_1$	10.6519	3.0476	3.5428	-0.1748	-17.0675
$f_2$	0.5668	1.0000	0.8512	-0.2440	-0.3908	$f_2$	3.0476	8.1245	10.6959	-2.2395	-19.6286
$f_3$	0.6880	0.8512	1.0000	-0.2653	-0.2860	$f_3$	3.5428	10.6959	76.0159	-34.4484	-55.8063
$f_4$	-0.2852	-0.2440	-0.2653	1.0000	-0.4751	$f_4$	-0.1748	-2.2395	-34.4484	62.4076	-25.5447
$f_5$	-0.4257	-0.3908	-0.2860	-0.4751	1.0000	$f_5$	-17.0675	-19.6286	-55.8063	-25.5447	118.0473

(c) Eigenvalues and respective eigenvectors of $R$						(d) Eigenvalues and respective eigenvectors of $K$					
	$e_1$	$e_2$	$e_3$	$e_4$	$e_5$		$e_1$	$e_2$	$e_3$	$e_4$	$e_5$
	0.5392	0.2938	0.0983	0.0516	0.0168		0.5875	0.3440	0.0493	0.0191	0.0000
	$V_1$	$V_2$	$V_3$	$V_4$	$V_5$		$V_1$	$V_2$	$V_3$	$V_4$	$V_5$
$f_1$	0.515	-0.018	0.706	-0.335	0.352	$f_1$	0.108	-0.042	-0.732	-0.500	0.447
$f_2$	0.548	-0.025	-0.527	0.325	0.563	$f_2$	0.144	0.009	-0.278	0.838	0.447
$f_3$	0.561	-0.092	-0.332	-0.459	-0.596	$f_3$	0.548	0.509	0.453	-0.188	0.447
$f_4$	-0.168	0.745	-0.246	-0.531	0.274	$f_4$	0.017	-0.798	0.391	-0.103	0.447
$f_5$	-0.303	-0.660	-0.231	-0.538	0.360	$f_5$	-0.817	0.321	0.166	-0.046	0.447

## 4.8 Working of the Proposed Algorithms when Applied to Data Sets Characterised by Noise

In this section it is intended to evaluate the ability of the proposed objective reduction algorithms in handling datasets with a disparate degree of redundancy, using the non-dominated solution sets  $\mathcal{N}_\epsilon$  and  $\mathcal{N}_{\mathcal{NS}}$ .

### 4.8.1 DTLZ5(3,5): L-PCA and NL-MVU-PCA based on $\mathcal{N}_{\mathcal{NS}}$

In this section the algorithms are demonstrated on DTLZ5(3,5) when applied to  $\mathcal{N}_{\mathcal{NS}}$ . The test problem DTLZ5(3,5) is again selected for demonstration purposes due to its moderate redundancy.

#### DTLZ5(3,5): L-PCA based on $\mathcal{N}_{\mathcal{NS}}$

The correlation matrix  $R$  is shown in Table 4.9a and the respective eigenvalues and eigenvectors in Table 4.9c. The analysis that follows is based on  $\mathcal{N}_{\mathcal{NS}}$  for DTLZ5(3,5) using L-PCA

Table 4.10: DTLZ5(3, 5): Iteration1 of L-PCA with  $\mathcal{N}_{\mathcal{NS}}$  (one run)

(a) Eigenvalue analysis							
PCA	Variance	Cumulative	Subsets Selected				
$(N_v)$	(%)	(%)	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
1	53.92	53.92			$f_3$	$f_4$	$f_5$
2	29.38	83.30	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
3	09.83	93.13	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
4	05.16	98.29		$f_2$			$f_5$
5	01.68	99.97	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$

(b) RCM analysis			
1) Correlation over whole set	Equation 4.1(i)	$\hat{\mathcal{S}}_1 = \hat{\mathcal{S}}_2 = \hat{\mathcal{S}}_3 = \{f_1, f_2, f_3\}$	
2) Correlation threshold: $T_{cor}$	Equation 4.2	$1.0 - 0.5392 \times (1.0 - 4/5) = 0.8921$	
3) Correlation meeting $T_{cor}$	Equation 4.1(ii)	$\mathcal{S}_1 = \mathcal{S}_2 = \mathcal{S}_3 = \emptyset$	

algorithm. The steps are:

1. The eigenvalue analysis in Table 4.10a shows that all the five principal components need to be considered to account for variance threshold  $\theta = 0.997$ . As a result, all five objectives would be considered as important, leading to  $\mathcal{F}_e = \{f_1, f_2, f_3, f_4, f_5\}$ .
2. The RCM analysis in Table 4.10b identifies the potential identical correlated objective sets as  $\hat{\mathcal{S}}_1 = \hat{\mathcal{S}}_2 = \hat{\mathcal{S}}_3 = \{f_1, f_2, f_3\}$ . However, since  $T_{cor} = 0.8921$  is bigger than  $R_{12} = 0.5668$ ,  $R_{13} = 0.6880$  and  $R_{23} = 0.8512$ , the objectives  $\{f_1, f_2, f_3\}$  cannot be considered identically correlated. As a result, the last step of the procedure (Section 4.1.5) is eliminated and it is declared  $\mathcal{F}_s = \mathcal{F}_e = \{f_1, f_2, f_3, f_4, f_5\}$ .

Following the described steps it is obvious that L-PCA, for this example, was unable to reduce any of the objectives from  $\mathcal{F}_t$ , hence,  $\mathcal{F}_{redn} = \emptyset$  and  $\mathcal{E}_t = 0$ .

#### DTLZ5(3,5): NL-MVU-PCA based on $\mathcal{N}_{\mathcal{NS}}$

The kernel matrix is shown in Table 4.9b and respective eigenvalues and eigenvectors in Table 4.9d. The analysis that follows is based on  $\mathcal{N}_{\mathcal{NS}}$  for DTLZ5(3, 5) using NL-MVU-PCA algorithm. The steps are:

1. The eigenvalue analysis in Table 4.11a shows that the first four principal components have accounted for variance greater than the threshold  $\theta = 0.997$ , leading to  $\mathcal{F}_e = \{f_1, f_2, f_3, f_4, f_5\}$ .
2. The RCM analysis in Table 4.11b captures  $\hat{\mathcal{S}}_1 = \hat{\mathcal{S}}_2 = \hat{\mathcal{S}}_3 = \{f_1, f_2, f_3\}$  as the potential correlated sets of objectives since they satisfy Equation 4.1(i). However, the condition from Equation 4.1(ii) is violated by pairs  $R_{12} = 0.5668$  and  $R_{13} = 0.6880$ . This did not happen with pairs  $R_{13} = 0.6880$  and  $R_{23} = 0.8512$ . After this procedure, the sets  $\mathcal{S}_2 = \mathcal{S}_3 = \{f_2, f_3\}$  are considered to be identically correlated sets.
3. The selection scheme in Table 4.11c selects the most significant objectives from  $\{f_2, f_3\}$ . As  $f_3$  is the objective with the highest selection criterion value, both  $\mathcal{S}_2$  and  $\mathcal{S}_3$  can be represented by  $f_3$  alone. This leads to  $\mathcal{F}_s = \{f_1, f_3, f_4, f_5\}$  and since objective  $f_2$  is deleted the error is  $\mathcal{E}_0 = c_2^M(1.0 - R_{23}) = 0.00439$  due to  $c_2^M = 0.029517$  and  $R_{23} = 0.8512$ .

In Iteration1 the algorithm could only reduce one objective, namely  $f_2$ , and to show the benefits of the iterative approach, Iteration2 is now described. The correlation and kernel matrix are shown in Tables 4.12a and 4.12b, respectively, and the eigenvalues and eigenvectors of  $K$  are shown in Table 4.12c. The remaining steps are as follows:

1. The eigenvalue analysis in Table 4.13a shows that the first three principal components have accounted for variance greater than threshold  $\theta = 0.997$ , leading to  $\mathcal{F}_e = \{F_1, F_2, F_3, F_4\}$  corresponding to objective set  $\{f_1, f_3, f_4, f_5\}$ .
2. The RCM analysis in Table 4.13b captures  $\hat{\mathcal{S}}_1 = \hat{\mathcal{S}}_2 = \{F_1, F_2\}$  as the potential correlated sets of objectives. Since the correlation  $R_{12} = 0.713$  is higher than  $T_{cor} = 0.6954$  the sets  $\mathcal{S}_1 = \mathcal{S}_2 = \{F_1, F_2\}$  are confirmed to be *identically* correlated.
3. The selection scheme in Table 4.13c eliminates  $F_1$  when compared with  $F_2$  which leads to  $\mathcal{F}_s = \{F_2, F_3, F_4\} = \{f_3, f_4, f_5\}$ .

Table 4.11: DTLZ5(3, 5): Iteration1 of NL-MVU-PCA with  $\mathcal{N}_{\mathcal{NS}}$  (one run)

(a) Eigenvalue analysis

PCA	Variance	Cumulative	Subsets Selected				
$(N_v)$	(%)	(%)	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
1	58.75	58.75	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
2	34.40	93.15		$f_2$	$f_3$	$f_4$	$f_5$
3	04.93	98.08	$f_1$		$f_3$	$f_4$	$f_5$
4	01.91	99.99	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$

(b) RCM analysis

1) Correlation over whole set	Equation 4.1(i)	$\hat{\mathcal{S}}_1 = \hat{\mathcal{S}}_2 = \hat{\mathcal{S}}_3 = \{f_1, f_2, f_3\}$
2) Correlation threshold: $T_{cor}$	Equation 4.2	$1.0 - 0.5875 \times (1.0 - 3/5) = 0.7649$
3) Correlation meeting $T_{cor}$	Equation 4.1(ii)	$\mathcal{S}_1 = \emptyset, \mathcal{S}_2 = \mathcal{S}_3 = \{f_2, f_3\}$

(c) Selection scheme

	$e_1=0.5875$	$e_2=0.3440$	$e_3=0.0493$	$e_4=0.0191$	
	$V_1$	$V_2$	$V_3$	$V_4$	$sc_i$
$f_1$	0.108	-0.042	-0.732	-0.500	0.124
$f_2$	0.144	-0.009	0.278	0.838	0.118
$f_3$	0.548	-0.509	-0.453	-0.188	0.523

Table 4.12: DTLZ5(3, 5): The  $R$  and  $K$  matrix with eigenvalues and eigenvectors of  $K$  Iteration2 of NL-MVU-PCA (one run)

(a) Correlation matrix $R$					(b) Kernel matrix $K$				
	$F_1 \equiv f_1$	$F_2 \equiv f_3$	$F_3 \equiv f_4$	$F_4 \equiv f_5$		$F_1 \equiv f_1$	$F_2 \equiv f_3$	$F_3 \equiv f_4$	$F_4 \equiv f_5$
$F_1$	1.0000	0.7513	-0.3201	-0.3855	$F_1$	8.3238	7.3581	-1.1085	-14.5733
$F_2$	0.7513	1.0000	-0.2823	-0.2227	$F_2$	7.3581	53.3221	-25.3016	-35.3786
$F_3$	-0.3201	-0.2823	1.0000	-0.5626	$F_3$	-1.1085	-25.3016	79.8718	-53.4616
$F_4$	-0.3855	-0.2227	-0.5626	1.0000	$F_4$	-14.5733	-35.3786	-53.4616	103.4136

(c) Eigenvalues and respective eigenvectors of $K$				
	$e_1$	$e_2$	$e_3$	$e_4$
	0.6091	0.3582	0.0326	0.0000
	$V_1$	$V_2$	$V_3$	$V_4$
$F_1$	0.087	0.126	0.852	0.500
$F_2$	0.155	0.742	-0.419	0.500
$F_3$	0.564	-0.603	-0.262	0.500
$F_4$	-0.806	-0.265	-0.172	0.500

Table 4.13: DTLZ5(3, 5): Iteration2 of NL-MVU-PCA with  $\mathcal{N}_{\mathcal{NS}}$  (one run)

(a) Eigenvalue analysis

PCA	Variance	Cumulative	Subsets Selected			
$(N_v)$	(%)	(%)	$F_1$	$F_2$	$F_3$	$F_4$
1	60.91	60.91	$F_1$	$F_2$	$F_3$	$F_4$
2	35.82	96.73		$F_2$	$F_3$	$F_4$
3	03.26	99.99	$F_1$	$F_2$	$F_3$	$F_4$

(b) RCM analysis

1) Correlation over whole set	Equation 4.1(i)	$\hat{\mathcal{S}}_1 = \hat{\mathcal{S}}_2 = \{F1, F2\}$
2) Correlation threshold: $T_{cor}$	Equation 4.2	$1.0 - 0.6091 \times (1.0 - 2/4) = 0.6954$
3) Correlation meeting $T_{cor}$	Equation 4.1(ii)	$\mathcal{S}_1 = \mathcal{S}_2 = \{F1, F2\}$

(c) Selection scheme

	$e_1=0.6091$	$e_2=0.3582$	$e_3=0.0326$	$e_4=0.0000$	
	$V_1$	$V_2$	$V_3$	$V_4$	$sc_i$
$F_1$	0.087	0.126	0.852	0.500	0.1259
$F_2$	0.155	0.742	-0.419	0.500	0.3742

4. Finally, the error computation for this reduction is  $\mathcal{E}_1 = c_1^M(1.0 - R_{12}) = 0.00845$  since  $c_2^M = 0.033984$  and  $R_{12} = 0.7513$ .

To conclude, the nonlinear NL-MVU-PCA procedure, could identify  $\mathcal{F}_s = \{f_3, f_4, f_5\}$  which is considered the true critical set for DTLZ5(3, 5) problem. The redundant set is then  $\mathcal{F}_{redn} = \{f_1, f_2\}$  and the error computation for this reduction after two iterations is given by:

$$\begin{aligned}
 \mathcal{E}_T &= \mathcal{E}_0 + \mathcal{E}_1(1 - \mathcal{E}_0) \\
 &= 0.00439 + 0.00845(1 - 0.00439) \\
 &= 0.01280 \text{ or } \approx 1.3\%.
 \end{aligned}$$

#### 4.8.2 DTLZ5(3,5): L-PCA and NL-MVU-PCA based on $\mathcal{N}_\epsilon$

In this section the algorithms are demonstrated on DTLZ5(3, 5) when applied to  $\mathcal{N}_\epsilon$ .

Table 4.14: DTLZ5(3, 5): The  $R$  and  $K$  matrix with their corresponding eigenvalues and eigenvectors, for  $\mathcal{N}_\epsilon$  (one run)

(a) Correlation matrix $R$						(b) Kernel matrix $K$					
	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$		$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
$f_1$	1.0000	0.9875	0.9736	-0.4310	-0.3384	$f_1$	2.4039	2.4928	4.1692	-4.4405	-4.6255
$f_2$	0.9875	1.0000	0.9644	-0.4496	-0.3040	$f_2$	2.4928	2.7527	4.4084	-5.0861	-4.5678
$f_3$	0.9736	0.9644	1.0000	-0.4689	-0.3229	$f_3$	4.1692	4.4084	8.1727	-9.0340	-7.7163
$f_4$	-0.4310	-0.4496	-0.4689	1.0000	-0.5136	$f_4$	-4.4405	-5.0861	-9.0340	27.0187	-8.4579
$f_5$	-0.3384	-0.3040	-0.3229	-0.5136	1.0000	$f_5$	-4.6255	-4.5678	-7.7163	-8.4579	25.3677

(c) Eigenvalues and respective eigenvectors of $R$						(d) Eigenvalues and respective eigenvectors of $K$					
	$e_1$	$e_2$	$e_3$	$e_4$	$e_5$		$e_1$	$e_2$	$e_3$	$e_4$	$e_5$
	0.6567	0.3007	0.0333	0.0069	0.0021		0.5322	0.4608	0.0057	0.0011	0.0000
	$V_1$	$V_2$	$V_3$	$V_4$	$V_5$		$V_1$	$V_2$	$V_3$	$V_4$	$V_5$
$f_1$	0.546	-0.056	-0.219	0.225	-0.775	$f_1$	0.059	0.269	-0.376	-0.764	-0.447
$f_2$	0.544	-0.030	-0.274	0.505	0.610	$f_2$	0.076	0.282	-0.553	0.639	-0.447
$f_3$	0.545	-0.030	-0.059	-0.820	0.165	$f_3$	0.142	0.490	0.730	0.086	-0.447
$f_4$	-0.285	-0.660	-0.686	-0.114	0.007	$f_4$	-0.821	-0.337	0.108	0.027	-0.447
$f_5$	-0.166	0.748	-0.635	-0.097	-0.020	$f_5$	-0.545	-0.704	0.091	0.011	-0.447

### DTLZ5(3,5): L-PCA based on $\mathcal{N}_\epsilon$

The correlation matrix  $R$  is shown in Table 4.14a and the respective eigenvalues and eigenvectors in Table 4.14c. The analysis that follows is based on  $\mathcal{N}_\epsilon$  for DTLZ5(3, 5) using L-PCA algorithm. The steps are:

1. The eigenvalue analysis in Table 4.15a shows that four principal components are necessary to account for the variance threshold  $\theta = 0.997$ . As a result, all five objective are picked leading to  $\mathcal{F}_e = \{f_1, f_2, f_3, f_4, f_5\}$ .
2. The RCM analysis in Table 4.15b identifies a potential identically correlated objective set  $\hat{\mathcal{S}}_1 = \{f_1, f_2, f_3\}$ , and since the correlation  $R_{12} = 0.9875$ ,  $R_{13} = 0.9736$ , and  $R_{23} = 0.9644$ , are higher than  $T_{cor} = 0.6059$ , the objectives  $\{f_1, f_2, f_3\}$  are indeed identically correlated.
3. The selection scheme in Table 4.15c selects the most significant objective from the objective set  $\{f_1, f_2, f_3\}$ . In that,  $f_1$  is picked which leads to  $\mathcal{F}_s = \{f_1, f_4, f_5\}$ , which is not the correct set of objectives.

Table 4.15: DTLZ5(3, 5): Iteration1 of L-PCA with  $\mathcal{N}_\epsilon$  (one run)

(a) Eigenvalue analysis						
PCA	Variance	Cumulative	Subsets Selected			
$(N_v)$	(%)	(%)	$f_1$	$f_2$	$f_3$	$f_4$
1	65.67	65.67	$f_1$			$f_4$
2	30.07	95.74	$f_1$	$f_2$	$f_3$	$f_4$
3	03.33	99.07				$f_4$
4	00.69	99.76	$f_1$	$f_2$	$f_3$	

(b) RCM analysis			
1) Correlation over whole set	Equation 4.1(i)	$\hat{\mathcal{S}}_1 = \{f_1, f_2, f_3\}$	
2) Correlation threshold: $T_{cor}$	Equation 4.2	$1.0 - 0.6567 \times (1.0 - 2/5) = 0.6059$	
3) Correlation meeting $T_{cor}$	Equation 4.1(ii)	$\mathcal{S}_1 = \{f_1, f_2, f_3\}$	

(c) Selection scheme					
	$e_1=0.6567$	$e_2=0.3007$	$e_3=0.0333$	$e_4=0.0069$	
	$V_1$	$V_2$	$V_3$	$V_4$	$sc_i$
$f_1$	0.546	-0.056	-0.219	0.225	0.3859
$f_2$	0.544	-0.030	-0.274	0.505	0.3805
$f_3$	0.545	-0.030	-0.059	-0.820	0.3749

4. The error computation due to this reduction is  $\mathcal{E}_t = c_2^M(1.0 - R_{21}) + c_3^M(1.0 - R_{31}) = 0.007766$  or 0.77%, where  $R_{21} = 0.9875$  and  $R_{31} = 0.9736$ . Since during the second iteration no further reduction was possible, hence, it is not shown.

Following the described steps it is obvious that L-PCA failed once more to identify the correct essential objective set, since  $\mathcal{F}_t = \{f_1, f_4, f_5\}$  and objective  $f_3$  is considered to be redundant erroneously, which incurs in error  $\mathcal{E} = 0.77\%$ .

#### DTLZ5(3,5): NL-MVU-PCA based on $\mathcal{N}_\epsilon$

The kernel matrix  $K$  is shown in Table 4.14b and respective eigenvalues and eigenvectors in Table 4.14d. The analysis that follows is based on  $\mathcal{N}_\epsilon$  for DTLZ5(3, 5) using NL-MVU-PCA algorithm. The steps are:

1. The eigenvalue analysis in Table 4.16a shows that only three principal components are



Table 4.16: DTLZ5(3, 5): Iteration1 of NL-MVU-PCA with  $\mathcal{N}_\epsilon$  (one run)

(a) Eigenvalue analysis							
PCA	Variance	Cumulative	Subsets Selected				
( $N_v$ )	(%)	(%)	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
1	53.22	53.22	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
2	46.08	99.30	$f_1$	$f_2$	$f_3$		$f_5$
3	00.57	99.87	$f_1$	$f_2$	$f_3$		

(b) RCM analysis			
1) Correlation over whole set	Equation 4.1(i)	$\hat{\mathcal{S}}_1 = \{f_1, f_2, f_3\}$	
2) Correlation threshold: $T_{cor}$	Equation 4.2	$1.0 - 0.5322 \times (1.0 - 2/5) = 0.6806$	
3) Correlation meeting $T_{cor}$	Equation 4.1(ii)	$\mathcal{S}_1 = \{f_1, f_2, f_3\}$	

(c) Selection scheme				
	$e_1=0.5322$	$e_2=0.4608$	$e_3=0.0057$	
	$V_1$	$V_2$	$V_3$	$sc_i$
$f_1$	0.059	0.269	-0.376	0.1580
$f_2$	0.076	0.282	-0.553	0.1743
$f_3$	0.142	0.490	0.730	0.3054

needed to account for the variance threshold  $\theta = 0.997$ . Eventually, this leads to all five objectives to be selected, hence,  $\mathcal{F}_e = \{f_1, f_2, f_3, f_4, f_5\}$ .

2. As previously shown by L-PCA, the RCM analysis in Table 4.16b identifies  $\hat{\mathcal{S}}_1 = \{f_1, f_2, f_3\}$  as a potential identically correlated set. Also, since  $T_{cor} = 0.6806$  is smaller than  $R_{12}$ ,  $R_{13}$  and  $R_{23}$  then the objectives in the set  $\{f_1, f_2, f_3\}$  are confirmed to be identically correlated.
3. The selection scheme in Table 4.16c as opposed to L-PCA selected  $f_3$  as the most important objective from the objective set  $\{f_1, f_2, f_3\}$ . This leads to  $\mathcal{F}_s = \{f_3, f_4, f_5\}$ .
4. The error computation due to this reduction is  $\mathcal{E}_t = c_1^M(1.0 - R_{13}) + c_2^M(1.0 - R_{23}) = 0.002455$  or 0.24%, where  $R_{13} = 0.9736$  and  $R_{23} = 0.9644$ . Since during the second iteration no further reduction was possible, hence, it is not shown.

Following the described steps NL-MVU-PCA could identify the correct essential objective set for DTLZ5(3, 5) as opposed to L-PCA, since  $\mathcal{F}_t = \{f_3, f_4, f_5\}$  which incurs in error  $\mathcal{E} = 0.24\%$ .

### 4.8.3 Discussion of Results

The previous reported results are interpreted in this section by considering the following two factors:

1. The first factor takes into account the strengths and limitations of NL-MVU-PCA and L-PCA. In respect to  $\mathcal{N}_{\mathcal{P}}$ , as mentioned in Section 4.7.2, L-PCA failure is attributed to principal components that have not captured the objectives variance correctly.
2. The second factor analyses the degree of conformance between the correlation–structure of the POF against that obtained by  $\mathcal{N}_{\epsilon}$  and  $\mathcal{N}_{\mathcal{NS}}$ . When the non-dominated set correlation–structure totally conforms with the POF then it can be referred to as perfect POF-representative. In order for that situation to arise the following three features are required:

- F1** The first feature relates to objectives that are in conflict with each of the remaining objectives, e.g., for  $\mathcal{N}_{\mathcal{P}}$  objectives  $f_4$  and  $f_5$  in DTLZ5(3, 5).
- F2** The second feature relates to objectives that are identically correlated with each other, e.g., for  $\mathcal{N}_{\mathcal{P}}$  objectives  $f_1$ ,  $f_2$  and  $f_3$  in DTLZ5(3, 5).
- F3** The third feature takes into consideration the strength of correlation among the different pairs of identically correlated objectives, e.g.,  $R_{12} = R_{13} = R_{23} = 1.0$  for  $\mathcal{N}_{\mathcal{P}}$  in DTLZ5(3, 5).

For a non-dominated set to be considered perfect POF-representative, the previous three features cannot be violated. However, for realistic non-dominated sets the presence of noise is a constant and a perfect strength of correlation may never be obtained, as evident in  $\mathcal{N}_{\epsilon}$

Table 4.17: DTLZ5(3, 5): POF-representation by  $\mathcal{N}_{\mathcal{P}}$ ,  $\mathcal{N}_{\epsilon}$  and  $\mathcal{N}_{\mathcal{NS}}$

Data	Source Table	Feature: F1	Feature: F2	Feature: F3		
		$f_4, f_5$ conflicting with all	$f_1 - f_2 - f_3$ identically correlated	$R_{12}$	$R_{13}$	$R_{23}$
$\mathcal{N}_{\mathcal{P}}$	4.6a	Yes	Yes	1.00	1.00	1.00
$\mathcal{N}_{\mathcal{NS}}$	4.9a	Yes	Yes	0.73	0.82	0.87
$\mathcal{N}_{\epsilon}$	4.14a	Yes	Yes	0.98	0.97	0.96

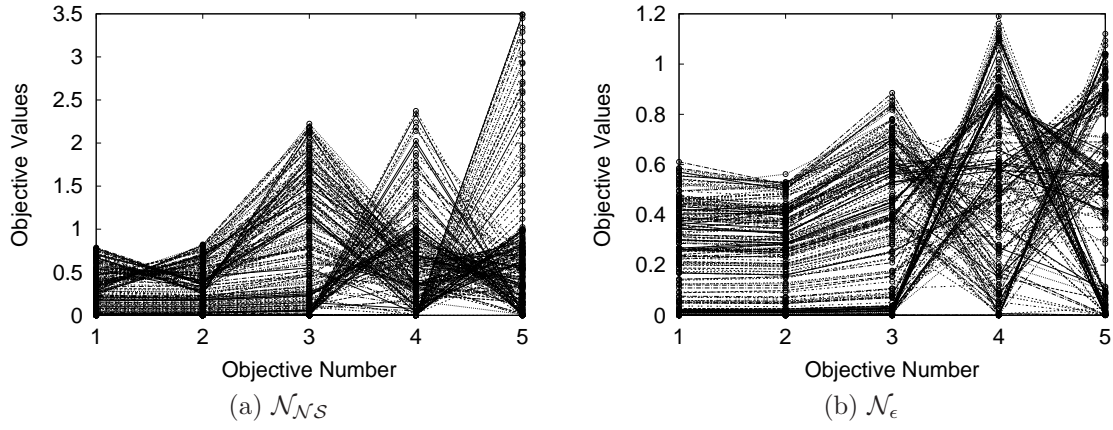


Figure 4.5: DTLZ5(3, 5): Parallel coordinate plots for  $\mathcal{N}_{\mathcal{NS}}$  and  $\mathcal{N}_{\epsilon}$  (one run).

and  $\mathcal{N}_{\mathcal{NS}}$ . It can be defined as good POF-representative when the strength of correlation of the identically correlated objectives are reasonably close to that captured on the POF.

Taking into account the previous discussion, Table 4.17 summarises for DTLZ5(3, 5) the POF-representation captured for  $\mathcal{N}_{\mathcal{P}}$ ,  $\mathcal{N}_{\epsilon}$  and  $\mathcal{N}_{\mathcal{NS}}$ . Both  $\mathcal{N}_{\epsilon}$  and  $\mathcal{N}_{\mathcal{NS}}$  have a good POF-representation since they respect features  $F_1$  and  $F_2$  and also because their correlation is reasonably close to that captured on the POF. In comparison, the POF-representation of  $\mathcal{N}_{\epsilon}$  is better than  $\mathcal{N}_{\mathcal{NS}}$  due to higher correlation strength in  $R_{12}$ ,  $R_{13}$  and  $R_{23}$ . The previous observation can be further validated with Figures 4.4 and 4.5 since the non-dominated solutions of  $\mathcal{N}_{\epsilon}$  conforms better with  $\mathcal{N}_{\mathcal{P}}$  and also because in  $\mathcal{N}_{\mathcal{NS}}$  some solutions are wrongly in conflict in respect to objectives  $\{f_1, f_2, f_3\}$ .

A more accurate NL-MVU-PCA over L-PCA and also a better POF-representation of  $\mathcal{N}_{\epsilon}$  over  $\mathcal{N}_{\mathcal{NS}}$  allows the results in Sections 4.8.1 and 4.8.2 to be explained and the following conclusions can be reached:

1. The performance of NL-MVU-PCA is higher than L-PCA for both  $\mathcal{N}_\epsilon$  and  $\mathcal{N}_{\mathcal{NS}}$ .
2. With  $\mathcal{N}_\epsilon$  both NL-MVU-PCA and L-PCA perform better than with  $\mathcal{N}_{\mathcal{NS}}$ .
3. The best performing configuration is NL-MVU-PCA with  $\mathcal{N}_\epsilon$  where only one iteration was needed.
4. The worst performing configuration is L-PCA with  $\mathcal{N}_{\mathcal{NS}}$  where no objective reduction took place.

#### 4.8.4 POF–representation versus POF–approximation

In the previous section the discussion of results and their performance is accessed based on their POF–representation (correlation–structure). In this section it is important to distinguish between POF–representation and POF–approximation.

A solution set can be evaluated based on the convergence and diversity towards the POF using quality indicators such as  $g$  and  $I_s$  (Section 4.4). A solution set nearly in the POF whose quality indicators are very close to their ideal values provides a good POF–approximation. As an example consider  $\mathcal{N}_{\mathcal{NS}}$  from Section 4.8.1 and  $\mathcal{N}_\epsilon$  from Section 4.8.2. Then the following measures arise:

Set	$g$	$I_s$
$\mathcal{N}_\epsilon$	0.07	1.48
$\mathcal{N}_{\mathcal{NS}}$	0.79	3.53
$\mathcal{N}_{\mathcal{P}}$	0.00	1.00

From obtained measures it can be said that  $\mathcal{N}_\epsilon$  provided a good POF–approximation but  $\mathcal{N}_{\mathcal{NS}}$  fail to do so. The previous observation can be endorsed by Figure 4.5a where  $\mathcal{N}_{\mathcal{NS}}$  POF–approximation is much worse than  $\mathcal{N}_\epsilon$ .

Having shown an example case of POF–approximation, it needs to be noted that  $\mathcal{N}_\epsilon$  provided a good POF–approximation and also a good POF–representation, while  $\mathcal{N}_{\mathcal{NS}}$  provided a good POF–representation but a bad POF–approximation. In that sense it can be

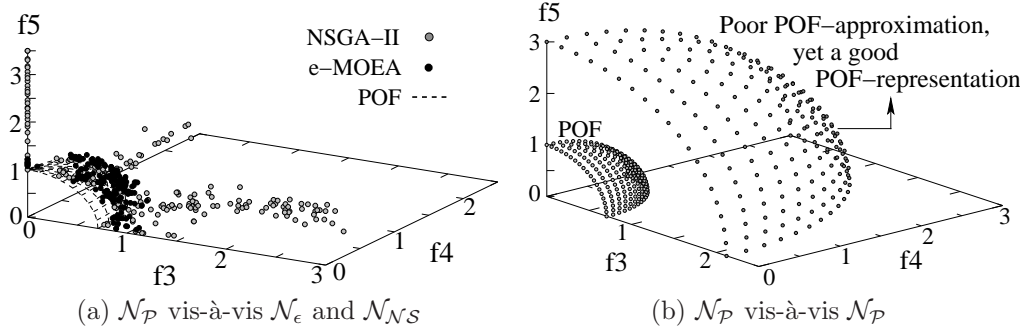


Figure 4.6: DTLZ5(3, 5): POF-representation versus POF-approximation.

inferred that *a good POF-approximation is a sufficient but not a necessary condition for a good POF-representation*. The previous inference holds true because a solution set with high convergence and good diversity is considered to be in the POF therefore the correlation-structure will also conform with that of the POF. However, this does not hold true the other way around, which means that *a good POF-representation is not necessary a good POF-approximation*. As an example, consider Figure 4.6b where the objectives values in  $\mathcal{N}_P$  are uniformly scaled up by a constant, and while the POF-representation of scaled  $\mathcal{N}_P$  is the same as that in the POF the POF-approximation is considered bad.

It can be concluded from the previous discussion that a solution set with good POF-approximation (ensures a good POF-representation) obtains the most accurate results from the algorithms. However, if the POF-approximation is considered to be bad then the results need to be interpreted based on their POF-representation.

#### 4.8.5 Benefit of Objective Reduction

In this section is it worth highlighting the benefits that objective reduction can offer. Consider DTLZ5(3, 5) as an example in Figure 4.7, where  $\mathcal{N}_{NS}$  is obtained from a run only with objectives in  $\mathcal{F}_T = \{f_3, f_4, f_5\}$ . For this case generating  $\mathcal{N}_{NS}$  with the reduced set of objectives ensures a totally converged non-dominated set of solutions at the expense of redundant objectives. However, if all the objectives are to be considered, convergence towards

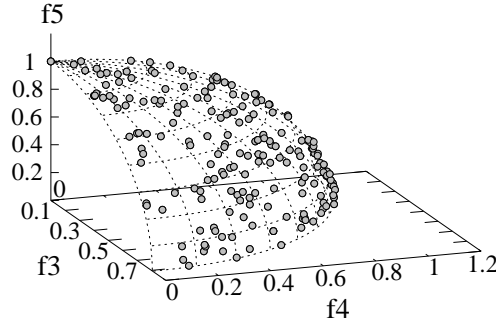


Figure 4.7: DTLZ5(3, 5): Illustrating that the  $\mathcal{N}_{NS}$  (one run) obtained for the reduced problem,  $\mathcal{F}_{\mathcal{T}} = \{f_3, f_4, f_5\}$ , conforms with the true POF.

POF suffers as shown in Figure 4.6a, due to a large amount of non-optimal solutions. This example highlights the importance of objective reduction in many-objective optimisation since the information captured regarding problem redundancy can be useful for an MOEA to converge towards the POF.

## 4.9 Experimental Results on a Wide Range of Redundant Test Problems and Comparison with DRP Based Algorithms

In this section DTLZ5( $I, M$ ) and WFG3( $M$ ) test problems are used to evaluate the performance of objective reduction algorithms when applied to non-dominated solutions sets with different qualities of POF-representation and POF-approximation. The proposed L-PCA and NL-MVU-PCA are compared against DRP-based greedy and exact algorithms using the  $\delta$ -MOSS perspective with  $\delta = 0$ , and the non-dominated solution sets considered are  $\mathcal{N}_{\mathcal{P}}$ ,  $\mathcal{N}_{NS}$ ,  $\mathcal{N}_{\epsilon}$  and  $\mathcal{N}_{\mathcal{R}}$ .

The analysis starts with  $\mathcal{N}_{\mathcal{P}}$  where all the objective reduction algorithms could accurately identify an  $\mathcal{F}_{\mathcal{T}}$ , for all problems as reported in Table 4.18. However, for most cases, L-PCA failed to pick the particular  $\mathcal{F}_{\mathcal{T}}$  whose objective account for higher variance, as points by Equations 4.7 and 4.9. This can be attributed to the limitations of PCA as mentioned in

Table 4.18: Redundant problems: Results based on  $\mathcal{N}_{\mathcal{P}}$  with  $\theta = 0.997$

All algorithms accurately identify  $\mathcal{F}_{\mathcal{T}} = \{f_k, f_{M-I+2}, \dots, f_M\}$  for DTLZ5( $I, M$ ), and  $\mathcal{F}_{\mathcal{T}} = \{f_k, f_M\}$  for WFG3( $M$ ) problems, where  $f_k$  varies as reported below

DTLZ5( $I, M$ )		Proposed approaches		DRP <sup>a</sup> [93, 2]: $\delta$ -MOSS, 0% Error	
$I$	$M$	NL-MVU-PCA	L-PCA	Greedy	Exact
2	05	$f_4$	$f_4$	$f_1$	$f_1$
2	20	$f_{19}$	$f_8$	$f_1$	$f_1$
2	50	$f_{49}$	$f_{17}$	$f_1$	$f_1$
3	05	$f_3$	$f_3$	$f_1$	$f_1$
3	20	$f_{18}$	$f_8$	$f_1$	$f_1$
5	10	$f_6$	$f_3$	$f_1$	$f_1$
5	20	$f_{16}$	$f_{13}$	$f_1$	$f_1$
7	10	$f_4$	$f_3$	$f_1$	$f_1$
7	20	$f_{14}$	$f_{13}$	$f_1$	$f_1$
WFG3	05	$f_4$	$f_2$	$f_1$	$f_1$
	15	$f_{14}$	$f_5$	$f_1$	$f_1$
	25	$f_{24}$	$f_{20}$	$f_1$	$f_1$

<sup>a</sup> The tabulated results are obtained using the source codes at: <http://www.tik.ee.ethz.ch/sop/download/supplementary/objectiveReduction/>. The same codes are used for the results presented later in Tables 4.19, 4.22 and 4.30.

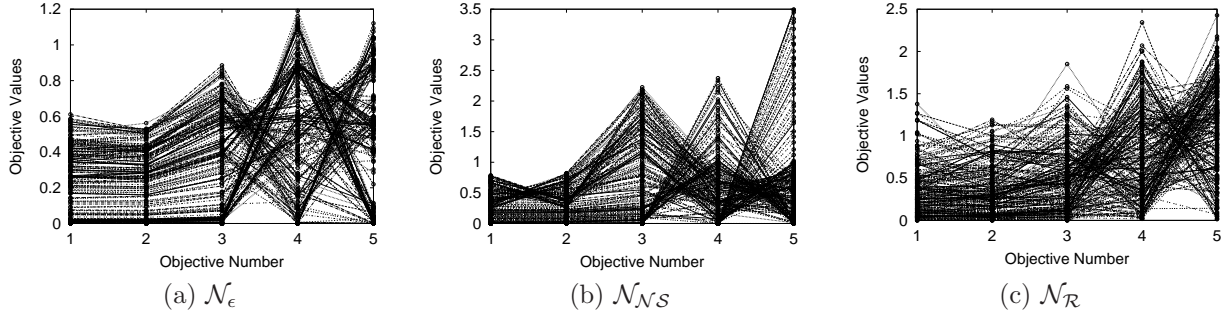


Figure 4.8: DTLZ5(3, 5): Parallel coordinate plots for  $\mathcal{N}_{\epsilon}$ ,  $\mathcal{N}_{NS}$  and  $\mathcal{N}_{\mathcal{R}}$  (one run).

#### Section 4.7.2.

Next, the results shown in Table 4.19 for  $\mathcal{N}_{\epsilon}$ ,  $\mathcal{N}_{NS}$  and  $\mathcal{N}_{\mathcal{R}}$ , show that the DRP-based algorithms failed to identify an essential objective set ( $\mathcal{F}_{\mathcal{T}}$ ), for all problems. This can be attributed to some nonoptimal solutions that show a conflict between objectives  $f_1$ - $f_2$ - $f_3$  for  $\mathcal{N}_{\epsilon}$ ,  $\mathcal{N}_{NS}$  and  $\mathcal{N}_{\mathcal{R}}$ , as perceived in Figures 4.8a, 4.8b and 4.8c, respectively. The nonoptimal solutions have different dominance relations from those solution on the POF. As a result when attempting to preserve the dominance relations of the global solution set the DRP-based algorithms fails to identify an  $\mathcal{F}_{\mathcal{T}}$ .





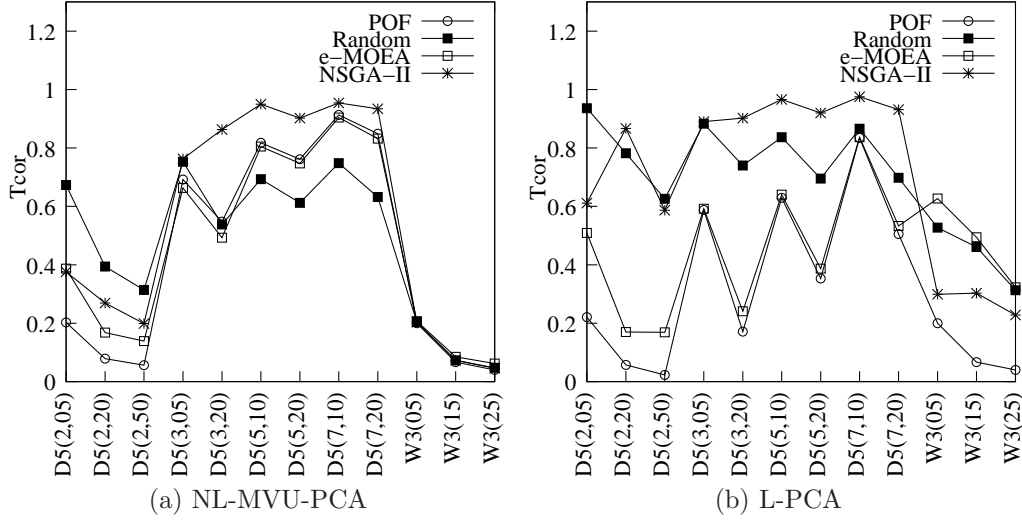


Figure 4.9: Redundant problems: On POF-representation by  $\mathcal{N}_P$ ,  $\mathcal{N}_R$ ,  $\mathcal{N}_\epsilon$  and  $\mathcal{N}_{NS}$ .

The following analysis is specifically for DTLZ5( $I, M$ ):

- For  $\mathcal{N}_\epsilon$ ,  $\mathcal{N}_{NS}$  and  $\mathcal{N}_R$  the performance of NL-MVU-PCA is superior to L-PCA. This could be attributed to the nonlinear unfolding process by NL-MVU-PCA that allows the algorithm to capture more accurate principal components than L-PCA, as argued in Section 4.7.2.
- The performance of NL-MVU-PCA and L-PCA is better with  $\mathcal{N}_\epsilon$  than with  $\mathcal{N}_{NS}$ , and the performance of the algorithms with  $\mathcal{N}_{NS}$  is better than with  $\mathcal{N}_R$ . This could be explained as  $T_{cor}$  values based on  $\mathcal{N}_\epsilon$  are closer to  $T_{cor}$  values based on  $\mathcal{N}_P$  than  $\mathcal{N}_{NS}$  and  $\mathcal{N}_R$ , as shown in Figure 4.9. This means that  $\mathcal{N}_\epsilon$  POF-representation is better than  $\mathcal{N}_{NS}$  and  $\mathcal{N}_R$ , which is further corroborated by the fact that  $\mathcal{N}_\epsilon$  has a good POF-approximation (shown in Table 4.20) which is considered a sufficient condition for a good POF-representation.

The following analysis is specifically for WFG3( $M$ ):

- The performance of NL-MVU-PCA is better than L-PCA for  $\mathcal{N}_\epsilon$ ,  $\mathcal{N}_{NS}$  and  $\mathcal{N}_R$ .
- The performance of NL-MVU-PCA and L-PCA is better with  $\mathcal{N}_{NS}$  than with  $\mathcal{N}_\epsilon$  and the worst performance is reported for  $\mathcal{N}_R$ . However, between  $\mathcal{N}_{NS}$  and  $\mathcal{N}_\epsilon$  the difference

Table 4.20: Redundant problems: on POF-approximation through convergence ( $g$ ) and diversity ( $I_s$ ) for  $\mathcal{N}_\epsilon$ ,  $\mathcal{N}_{\mathcal{NS}}$  and  $\mathcal{N}_{\mathcal{R}}$

The mean ( $\mu$ ) and standard deviation ( $\sigma$ ) are averaged over 20 runs							
Test Problems		Convergence ( $g$ )			Diversity ( $I_s$ )		
DTLZ5( $I, M$ )		$\mathcal{N}_\epsilon$	$\mathcal{N}_{\mathcal{NS}}$	$\mathcal{N}_{\mathcal{R}}$	$\mathcal{N}_\epsilon$	$\mathcal{N}_{\mathcal{NS}}$	$\mathcal{N}_{\mathcal{R}}$
$I$	$M$	( $\mu \pm \sigma$ )	( $\mu \pm \sigma$ )	( $\mu \pm \sigma$ )	( $\mu \pm \sigma$ )	( $\mu \pm \sigma$ )	( $\mu \pm \sigma$ )
2	05	$0.15 \pm 0.09$	$0.48 \pm 0.63$	$0.83 \pm 0.02$	$1.93 \pm 0.01$	$4.20 \pm 0.17$	$3.89 \pm 0.12$
2	20	$0.20 \pm 0.07$	$2.10 \pm 0.58$	$0.83 \pm 0.02$	$1.91 \pm 0.02$	$8.06 \pm 0.17$	$4.05 \pm 0.13$
2	50	$0.23 \pm 0.08$	$2.34 \pm 0.34$	$0.82 \pm 0.01$	$1.99 \pm 0.02$	$8.18 \pm 0.22$	$4.02 \pm 0.12$
3	05	$0.08 \pm 0.04$	$0.70 \pm 0.61$	$0.83 \pm 0.02$	$1.48 \pm 0.00$	$3.54 \pm 0.02$	$2.92 \pm 0.09$
3	20	$0.17 \pm 0.07$	$2.25 \pm 0.03$	$0.83 \pm 0.02$	$1.59 \pm 0.02$	$6.67 \pm 0.01$	$3.03 \pm 0.10$
5	10	$0.14 \pm 0.07$	$2.06 \pm 0.34$	$0.83 \pm 0.02$	$1.38 \pm 0.02$	$4.29 \pm 0.02$	$2.29 \pm 0.09$
5	20	$0.15 \pm 0.07$	$2.25 \pm 0.31$	$0.83 \pm 0.02$	$1.37 \pm 0.00$	$5.00 \pm 0.01$	$2.30 \pm 0.08$
7	10	$0.16 \pm 0.07$	$1.99 \pm 0.38$	$0.83 \pm 0.02$	$1.27 \pm 0.01$	$3.30 \pm 0.01$	$1.93 \pm 0.08$
7	20	$0.16 \pm 0.08$	$2.17 \pm 0.38$	$0.83 \pm 0.02$	$1.28 \pm 0.00$	$3.94 \pm 0.02$	$1.93 \pm 0.08$

is only marginal in terms of average dimension of the POF, as shown in Table 4.19b.

This implies that the POF-representation of  $\mathcal{N}_{\mathcal{NS}}$  is better than  $\mathcal{N}_\epsilon$ . This is validated by the fact that  $T_{cor}$  values based on  $\mathcal{N}_{\mathcal{NS}}$  are closer to  $T_{cor}$  values based on  $\mathcal{N}_{\mathcal{P}}$  than  $\mathcal{N}_\epsilon$  (Figure 4.9). Also, the performance of L-PCA as against NL-MVU-PCA is much worse for  $\mathcal{N}_\epsilon$  because the  $T_{cor}$  values based on  $\mathcal{N}_\epsilon$  are much further away than those based on  $\mathcal{N}_{\mathcal{NS}}$ .

Based on the previous discussion the following observations can be made:

1. Considering the role of  $T_{cor}$ , a low and high  $T_{cor}$  are desirable for problems with high and low redundancy, respectively. This is corroborated by the fact that  $T_{cor}$  is low when  $I$  is also low, and  $T_{cor}$  is high when  $I$  is also high, for all problems as shown in Figure 4.9.
2. The cases that require more than one iteration from the framework are when: (i) the correlations between identically correlated objectives could not be captured or (ii) the correlation strength between identically correlated objectives is lower than the derived  $T_{cor}$ . As a result, in one iteration only some of the redundant objectives are eliminated, and the remainder in subsequent generations.

Table 4.21: Errors ( $\mathcal{E}_T$ ) associated with NL-MVU-PCA and L-PCA when applied to  $\mathcal{N}_\epsilon$ ,  $\mathcal{N}_{\mathcal{N}_S}$  and  $\mathcal{N}_{\mathcal{R}}$  for redundant test problems, with  $\theta = 0.997$

The mean ( $\mu$ ) and standard deviation ( $\sigma$ ) is obtained for 20 random runs							
Test problems		NL-MVU-PCA			L-PCA		
DTLZ5( $I, M$ )		$\mathcal{N}_\epsilon$	$\mathcal{N}_{\mathcal{N}_S}$	$\mathcal{N}_{\mathcal{R}}$	$\mathcal{N}_\epsilon$	$\mathcal{N}_{\mathcal{N}_S}$	$\mathcal{N}_{\mathcal{R}}$
$I$	$M$	( $\mu \pm \sigma$ )	( $\mu \pm \sigma$ )	( $\mu \pm \sigma$ )	( $\mu \pm \sigma$ )	( $\mu \pm \sigma$ )	( $\mu \pm \sigma$ )
2	05	0.00747 $\pm$ 0.00435	0.01541 $\pm$ 0.01271	0.012662 $\pm$ 0.004914	0.05032 $\pm$ 0.02293	0.03306 $\pm$ 0.02591	0.000000 $\pm$ 0.000000
2	20	0.00956 $\pm$ 0.00481	0.02437 $\pm$ 0.02425	0.028254 $\pm$ 0.013311	0.05882 $\pm$ 0.02542	0.01173 $\pm$ 0.01836	0.057305 $\pm$ 0.015715
2	50	0.01565 $\pm$ 0.01095	0.06615 $\pm$ 0.02003	0.023517 $\pm$ 0.017610	0.10391 $\pm$ 0.04306	0.21139 $\pm$ 0.05023	0.171296 $\pm$ 0.032844
3	05	0.00113 $\pm$ 0.00085	0.01382 $\pm$ 0.00566	0.011931 $\pm$ 0.006499	0.00462 $\pm$ 0.00289	0.00000 $\pm$ 0.00000	0.000000 $\pm$ 0.000000
3	20	0.00547 $\pm$ 0.00394	0.00020 $\pm$ 0.00054	0.016962 $\pm$ 0.007100	0.05459 $\pm$ 0.02273	0.00044 $\pm$ 0.00113	0.098825 $\pm$ 0.018461
5	10	0.00159 $\pm$ 0.00092	0.00000 $\pm$ 0.00000	0.006346 $\pm$ 0.001362	0.01454 $\pm$ 0.00624	0.00000 $\pm$ 0.00000	0.029452 $\pm$ 0.014305
5	20	0.00258 $\pm$ 0.00127	0.00000 $\pm$ 0.00000	0.007345 $\pm$ 0.002466	0.02955 $\pm$ 0.01351	0.00000 $\pm$ 0.00000	0.107597 $\pm$ 0.015475
7	10	0.00179 $\pm$ 0.00086	0.00000 $\pm$ 0.00000	0.002101 $\pm$ 0.000631	0.01052 $\pm$ 0.00465	0.00000 $\pm$ 0.00000	0.017608 $\pm$ 0.006579
7	20	0.00218 $\pm$ 0.00077	0.00000 $\pm$ 0.00000	0.003816 $\pm$ 0.001222	0.03391 $\pm$ 0.01248	0.00000 $\pm$ 0.00000	0.081754 $\pm$ 0.014260
WFG3	05	0.00725 $\pm$ 0.00114	0.00261 $\pm$ 0.00074	0.004448 $\pm$ 0.000639	0.09986 $\pm$ 0.04019	0.03722 $\pm$ 0.01396	0.084825 $\pm$ 0.009404
	15	0.00962 $\pm$ 0.00296	0.00293 $\pm$ 0.00112	0.001856 $\pm$ 0.001463	0.24881 $\pm$ 0.18592	0.21041 $\pm$ 0.05908	0.280606 $\pm$ 0.234138
	25	0.01002 $\pm$ 0.00511	0.00266 $\pm$ 0.00177	0.001787 $\pm$ 0.001589	0.13044 $\pm$ 0.09724	0.19117 $\pm$ 0.05063	0.091783 $\pm$ 0.083467

3. Based on the previous arguments the interpretation of the error reported in Table 4.21

by the objective reduction operations needs to be interpreted in the wake of: (a) the accuracy of the objective reduction algorithms and (b) the quality of the POF-representation inherit in the non-dominated solution sets. In that:

- The error reported by NL-MVU-PCA is less than L-PCA, for  $\mathcal{N}_\epsilon$ ,  $\mathcal{N}_{\mathcal{N}_S}$  and  $\mathcal{N}_{\mathcal{R}}$ . This holds true when both NL-MVU-PCA and L-PCA could identify an  $\mathcal{F}_T$ . For cases where L-PCA error is lower than NL-MVU-PCA the results can be explained by the fact that L-PCA could not reduce the objective set towards  $\mathcal{F}_T$ , as reported in Table 4.19a.
- The error reported based on  $\mathcal{N}_\epsilon$  is less than the error based on  $\mathcal{N}_{\mathcal{N}_S}$ , and the error based on  $\mathcal{N}_{\mathcal{N}_S}$  is less than the error based on  $\mathcal{N}_{\mathcal{R}}$ . This can be attributed to a better POF-representation by  $\mathcal{N}_\epsilon$  when compared with  $\mathcal{N}_{\mathcal{N}_S}$ , and by a better POF-representation by  $\mathcal{N}_{\mathcal{N}_S}$  when compared with  $\mathcal{N}_{\mathcal{R}}$ . This is also corroborated by the fact that the POF-approximation of  $\mathcal{N}_\epsilon$  is considered to be good which is a sufficient condition for a good POF-representation. In some situations the error based on  $\mathcal{N}_{\mathcal{N}_S}$  was lower than  $\mathcal{N}_\epsilon$  for L-PCA. This could also be attributed to the limitations of L-PCA that influence the captured principal components.

## 4.10 Experimental Results on a Wide Range of Non-Redundant Test Problems and Comparison with DRP Based Algorithms

In this section the test problems DTLZ1( $M$ ), DTLZ2( $M$ ), DTLZ3( $M$ ) and DTLZ4( $M$ ) are used to evaluate the performance of objective reduction algorithms when applied to non-dominated solutions sets with different qualities of POF-representation and POF-approximation. As in the previous section, the proposed L-PCA and NL-MVU-PCA are compared against DRP-based greedy and exact algorithms using the  $\delta$ -MOSS perspective with  $\delta = 0$ , and the non-dominated solution sets considered are  $\mathcal{N}_{\mathcal{P}}$ ,  $\mathcal{N}_{\epsilon}$ ,  $\mathcal{N}_{\mathcal{NS}}$  and  $\mathcal{N}_{\mathcal{R}}$ .

The analysis starts with  $\mathcal{N}_{\mathcal{P}}$ , where all the objective reduction algorithms could accurately identify any absence of redundancy leading to  $m = M$  which implies  $\mathcal{F}_{\mathcal{T}} = \{f_1, \dots, f_M\}$ , for all problems.

Next, the analysis regarding the results shown in Table 4.22 for  $\mathcal{N}_{\epsilon}$ ,  $\mathcal{N}_{\mathcal{NS}}$  and  $\mathcal{N}_{\mathcal{R}}$ , is as follows:

- The DRP-based algorithms have failed to identify  $m = M$ , with the exception of the five-objective instance problems.
- The NL-MVU-PCA and L-PCA could identify  $m = M$  for all problems but, a lower performance is reported for  $\mathcal{N}_{\epsilon}$  when compared with  $\mathcal{N}_{\mathcal{NS}}$ , and the lowest performance is reported for  $\mathcal{N}_{\mathcal{R}}$ .

To understand the divergence between the reported performances, the following analysis needs to be interpreted by considering the difference between POF-approximation and POF-representation (mentioned in Section 4.8.4) inherent in  $\mathcal{N}_{\epsilon}$ ,  $\mathcal{N}_{\mathcal{NS}}$  and  $\mathcal{N}_{\mathcal{R}}$ :

1. The deviation between the  $T_{cor}$  values based on  $\mathcal{N}_{\mathcal{NS}}$  and  $\mathcal{N}_{\mathcal{P}}$  is very low as shown in Figure 4.10. This means that the POF-representation of  $\mathcal{N}_{\mathcal{NS}}$  is considered to be good.

Table 4.22: Non-redundant problems: Results based on  $\mathcal{N}_\epsilon$ ,  $\mathcal{N}_{\mathcal{NS}}$  and  $\mathcal{N}_{\mathcal{R}}$  with  $\theta = 0.997$

For brevity,  $P(M)$  represents an  $M$ -objective DTLZP problem. The entries are formatted as  $n(\mu \pm \sigma)$ , where  $n$  indicates the number of times that  $\mathcal{F}_{\mathcal{T}}$  is accurately identified, in 20 runs, while  $\mu$  and  $\sigma$  represent the mean and standard deviation of the number of essential objectives identified, in 20 runs

(a) Proposed Algorithms

$P(M)$	NL-MVU-PCA			L-PCA		
	$\mathcal{N}_\epsilon$	$\mathcal{N}_{\mathcal{NS}}$	$\mathcal{N}_{\mathcal{R}}$	$\mathcal{N}_\epsilon$	$\mathcal{N}_{\mathcal{NS}}$	$\mathcal{N}_{\mathcal{R}}$
1(05)	20(05.0±0.0)	20(05.0±0.0)	20(05.0±0.0)	20(05.0±0.0)	20(05.0±0.0)	20(05.0±0.0)
1(15)	06(14.1±0.7)	20(15.0±0.0)	02(12.4±1.6)	14(14.7±0.4)	20(15.0±0.0)	18(14.6±0.8)
1(25)	00(20.5±2.2)	20(25.0±0.0)	00(17.4±2.3)	03(22.2±2.1)	20(25.0±0.0)	06(23.2±1.7)
2(05)	20(05.0±0.0)	20(05.0±0.0)	20(05.0±0.0)	20(04.9±0.2)	20(05.0±0.0)	20(05.0±0.0)
2(15)	20(15.0±0.0)	20(15.0±0.0)	06(13.7±1.1)	20(15.0±0.0)	20(15.0±0.0)	19(14.9±0.2)
2(25)	16(24.7±0.5)	20(25.0±0.0)	00(19.0±1.8)	17(24.7±0.5)	20(25.0±0.0)	06(23.3±1.7)
3(05)	20(05.0±0.0)	20(05.0±0.0)	20(05.0±0.0)	20(05.0±0.0)	20(05.0±0.0)	20(05.0±0.0)
3(15)	11(14.3±0.9)	20(15.0±0.0)	05(13.6±1.1)	20(14.8±0.4)	20(15.0±0.0)	20(15.0±0.0)
3(25)	00(20.9±2.2)	20(25.0±0.0)	00(18.8±1.7)	05(22.0±2.6)	20(25.0±0.0)	06(23.1±2.0)
4(05)	20(05.0±0.0)	20(05.0±0.0)	20(05.0±0.0)	20(05.0±0.0)	20(05.0±0.0)	20(05.0±0.0)
4(15)	20(15.0±0.0)	20(15.0±0.0)	20(15.0±0.0)	20(15.0±0.0)	20(15.0±0.0)	20(15.0±0.0)
4(25)	20(25.0±0.0)	20(25.0±0.0)	20(25.0±0.0)	20(25.0±0.0)	20(25.0±0.0)	20(25.0±0.0)

(b) DRP [93, 2]:  $\delta$ -MOSS, 0% Error

$P(M)$	Greedy Approach			Exact Approach		
	$\mathcal{N}_\epsilon$	$\mathcal{N}_{\mathcal{NS}}$	$\mathcal{N}_{\mathcal{R}}$	$\mathcal{N}_\epsilon$	$\mathcal{N}_{\mathcal{NS}}$	$\mathcal{N}_{\mathcal{R}}$
1(05)	20(05.0±0.0)	20(05.0±0.0)	20(05.0±0.0)	20(05.0±0.0)	20(05.0±0.0)	20(05.0±0.0)
1(15)	08(13.8±1.0)	19(14.9±0.2)	00(11.5±0.9)	07(13.7±1.0)	19(14.9±0.2)	00(10.5±0.5)
1(25)	00(16.3±1.6)	02(23.1±1.2)	00(11.9±1.0)	00(16.2±1.4)	03(23.3±1.0)	00(11.6±0.9)
2(05)	20(05.0±0.0)	20(05.0±0.0)	20(05.0±0.0)	20(05.0±0.0)	20(05.0±0.0)	20(05.0±0.0)
2(15)	01(11.4±0.9)	17(14.9±0.4)	00(11.0±0.6)	00(10.8±0.7)	17(14.9±0.4)	00(10.7±0.7)
2(25)	00(12.4±1.0)	01(21.5±1.2)	00(11.3±0.8)	00(12.1±1.0)	01(21.7±1.1)	00(11.2±0.9)
3(05)	20(05.0±0.0)	20(05.0±0.0)	20(05.0±0.0)	20(05.0±0.0)	20(05.0±0.0)	20(05.0±0.0)
3(15)	08(13.1±0.9)	18(14.9±0.3)	00(12.9±1.0)	02(12.7±0.9)	18(14.9±0.3)	00(12.3±0.8)
3(25)	00(15.5±1.5)	06(24.0±0.9)	00(13.8±1.9)	00(14.8±1.2)	06(24.2±0.7)	00(13.6±1.7)
4(05)	14(04.7±0.4)	20(05.0±0.0)	20(05.0±0.0)	20(05.0±0.0)	20(05.0±0.0)	20(05.0±0.0)
4(15)	00(11.6±0.8)	00(12.1±0.9)	00(11.2±0.7)	00(11.0±0.6)	00(11.8±0.7)	00(11.5±1.0)
4(25)	00(11.4±0.8)	00(11.3±0.6)	00(12.0±0.8)	00(11.2±0.7)	00(11.3±0.6)	00(12.2±0.7)

As an example, consider DTLZ1(15) and DTLZ2(15) where conflicts between objectives have been captured as shown in Figures 4.11g and 4.11h, respectively. This implies a correlation-structure equal to the one on the true POF. Notably, the objectives in  $\mathcal{N}_{NS}$  have a higher objective scale than  $\mathcal{N}_{\mathcal{P}}$ , as shown in Figure 4.11. It implies that  $\mathcal{N}_{NS}$  POF-approximation is poor, as confirmed by Table 4.23. This corroborates the statement “a good POF-representation is not necessarily a good POF-approximation” as mentioned in Section 4.8.4, which explains the accuracy of the results.

2. For  $\mathcal{N}_{\epsilon}$ , NL-MVU-PCA and L-PCA have performed well for DTLZ2( $M$ ) and DTLZ4( $M$ ) problems. All the problem instances are accurately solved with the exception of DTLZ2(25). This can be explained by the low deviation between the  $T_{cor}$  values based on  $\mathcal{N}_{\epsilon}$  and  $\mathcal{N}_{\mathcal{P}}$ , which implies a good POF-representation. As an example, consider DTLZ2(15) where the conflict between objectives have been captured, as shown in Figure 4.11e. This implies a POF-representation equal to the one captured on the POF. Notably, the POF-approximation is considered reasonably good, which is a sufficient condition for a good POF-representation and that explains the obtained results. The only exception in this analysis is DTLZ2(25) which can be attributed to a partial convergence towards the POF since  $I_s \ll 1$  (diversity measure reported in Table 4.23), leading to a distorted correlation-structure. The reason for a partial convergence can be attributed to the inefficacy of the  $\epsilon$ -dominance strategy, found in  $\epsilon$ -MOEA, in diversity maintenance. This might result in a weak dispersion of solutions across the objective space as shown in Figures 4.11d, 4.11e and 4.11f, and in some cases this might mislead NL-MVU-PCA in identifying the true problem dimensionality. Note that, this is not the case with NSGA-II since the diversity operator, namely crowding distance, is able to preserve the diversity of the population as shown across all the objectives in Figures 4.11g, 4.11h and 4.11i.

3. For  $\mathcal{N}_{\epsilon}$ , NL-MVU-PCA and L-PCA have performed poorly for DTLZ1( $M$ ) and DTLZ3( $M$ )

problems. This is explained by the poor POF-representation of  $\mathcal{N}_\epsilon$  since the deviation of the  $T_{cor}$  values based on  $\mathcal{N}_\epsilon$  and  $\mathcal{N}_\mathcal{P}$  is high, as shown in Figure 4.10. An in depth analysis leads to the following two points:

- (a) The performance reported by the mean of the number of essential objectives identified ( $\mu$ ), is not as poor as the number of times that  $\mathcal{F}_\mathcal{T}$  is accurately identified ( $n$ ). As an example, NL-MVU-PCA applied to DTLZ1(15) gives  $n = 6$  and  $\mu = 14.1$ , while L-PCA gives  $n = 14$  and  $\mu = 14.7$ . Notably, these results are still better than DRP-based algorithms where  $\mu = 13.8$  and  $\mu = 13.7$  for greedy and exact, respectively.
  - (b) The lower the deviation between  $T_{cor}$  values based on  $\mathcal{N}_\epsilon$  and  $\mathcal{N}_\mathcal{P}$  the better the performance obtained. For example, consider  $M = 25$  and for both NL-MVU-PCA and L-PCA the deviation between the  $T_{cor}$  values reduces in the following order: DTLZ4, DTLZ2, DTLZ3 and DTLZ1. Notably, the performance measured by  $\mu$  also follows the same trend.
4. For  $\mathcal{N}_\mathcal{R}$ , NL-MVU-PCA and L-PCA have performed poorly for the majority of the problem instances and the only exception is DTLZ4. This is explained by the poor POF-representation of  $\mathcal{N}_\mathcal{R}$  since the deviation of the  $T_{cor}$  values based on  $\mathcal{N}_\mathcal{R}$  and  $\mathcal{N}_\mathcal{P}$  is high, as shown in Figure 4.10. As an example consider DTLZ2(15) where there is lack of conflict between some objectives, as shown in Figure 4.11k. This indicates that the POF-representation of  $\mathcal{N}_\mathcal{R}$  is very different from the one on the POF which explains the poor results. For DTLZ4 consider  $M = 15$  as example where the conflicts between the objectives have been captured, as shown in Figure 4.11l. Additionally, the low deviation between the  $T_{cor}$  values based on  $\mathcal{N}_\mathcal{R}$  and  $\mathcal{N}_\mathcal{P}$ , as shown in Figure 4.10, implies that the POF-representation of  $\mathcal{N}_\mathcal{R}$  is close to the one on the POF. This explains the accuracy of the results based on  $\mathcal{N}_\mathcal{R}$  for the DTLZ4 problem.

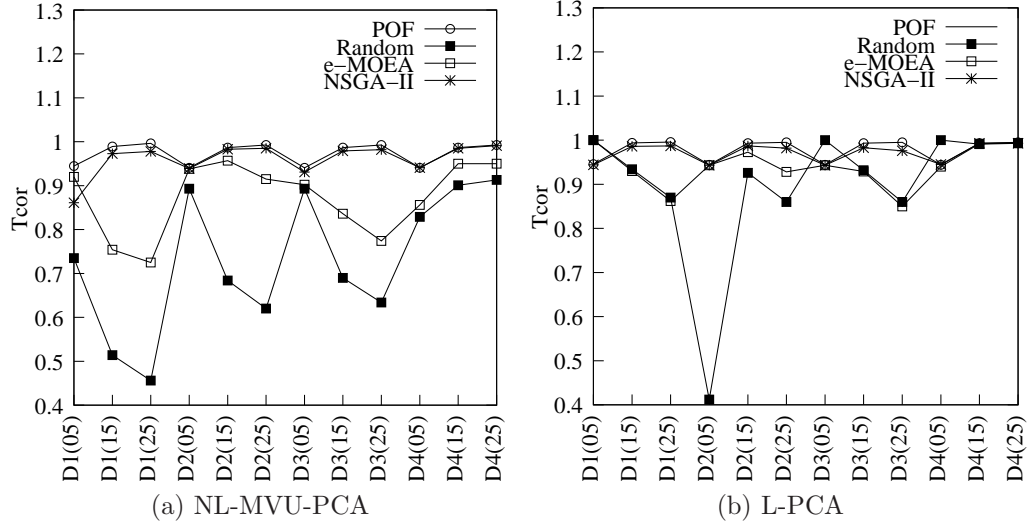


Figure 4.10: Non-redundant problems: On POF-representation by  $\mathcal{N}_P$ ,  $\mathcal{N}_R$ ,  $\mathcal{N}_\epsilon$  and  $\mathcal{N}_{NS}$ .

Table 4.23: Non-redundant problems: on POF-approximation through convergence ( $g$ ) and diversity ( $I_s$ ) for  $\mathcal{N}_\epsilon$ ,  $\mathcal{N}_{NS}$  and  $\mathcal{N}_R$

The mean ( $\mu$ ) and standard deviation ( $\sigma$ ) are averaged over 20 runs						
$P(M)$	Convergence ( $g$ )			Diversity ( $I_s$ )		
	$\mathcal{N}_\epsilon$ ( $\mu \pm \sigma$ )	$\mathcal{N}_{NS}$ ( $\mu \pm \sigma$ )	$\mathcal{N}_R$ ( $\mu \pm \sigma$ )	$\mathcal{N}_\epsilon$ ( $\mu \pm \sigma$ )	$\mathcal{N}_{NS}$ ( $\mu \pm \sigma$ )	$\mathcal{N}_R$ ( $\mu \pm \sigma$ )
1(05)	170 $\pm$ 129	410 $\pm$ 141	543 $\pm$ 11	480 $\pm$ 0	539 $\pm$ 6	558 $\pm$ 30
1(15)	313 $\pm$ 154	948 $\pm$ 165	541 $\pm$ 13	321 $\pm$ 4	507 $\pm$ 2	310 $\pm$ 16
1(25)	346 $\pm$ 152	944 $\pm$ 207	537 $\pm$ 12	269 $\pm$ 1	331 $\pm$ 6	244 $\pm$ 15
2(05)	0.09 $\pm$ 0.05	0.11 $\pm$ 0.06	0.83 $\pm$ 0.02	1.16 $\pm$ 0.00	1.08 $\pm$ 0.02	2.00 $\pm$ 0.06
2(15)	0.23 $\pm$ 0.11	2.08 $\pm$ 0.45	0.83 $\pm$ 0.02	0.99 $\pm$ 0.01	2.41 $\pm$ 0.00	1.23 $\pm$ 0.04
2(25)	0.22 $\pm$ 0.11	2.12 $\pm$ 0.49	0.83 $\pm$ 0.01	0.77 $\pm$ 0.01	2.10 $\pm$ 0.00	0.96 $\pm$ 0.04
3(05)	709 $\pm$ 223	0738 $\pm$ 220	1081 $\pm$ 19	1168 $\pm$ 5	1047 $\pm$ 18	1288 $\pm$ 50
3(15)	903 $\pm$ 225	1733 $\pm$ 356	1081 $\pm$ 19	0880 $\pm$ 7	1358 $\pm$ 08	808 $\pm$ 31
3(25)	908 $\pm$ 220	1808 $\pm$ 400	1083 $\pm$ 13	0669 $\pm$ 4	1187 $\pm$ 13	616 $\pm$ 32
4(05)	0.12 $\pm$ 0.06	0.13 $\pm$ 0.06	0.83 $\pm$ 0.02	1.13 $\pm$ 0.05	1.14 $\pm$ 0.00	1.72 $\pm$ 0.19
4(15)	0.22 $\pm$ 0.09	2.20 $\pm$ 0.15	0.83 $\pm$ 0.02	1.21 $\pm$ 0.00	2.80 $\pm$ 0.01	1.61 $\pm$ 0.11
4(25)	0.29 $\pm$ 0.13	2.18 $\pm$ 0.15	0.83 $\pm$ 0.01	1.24 $\pm$ 0.01	2.63 $\pm$ 0.00	1.60 $\pm$ 0.09



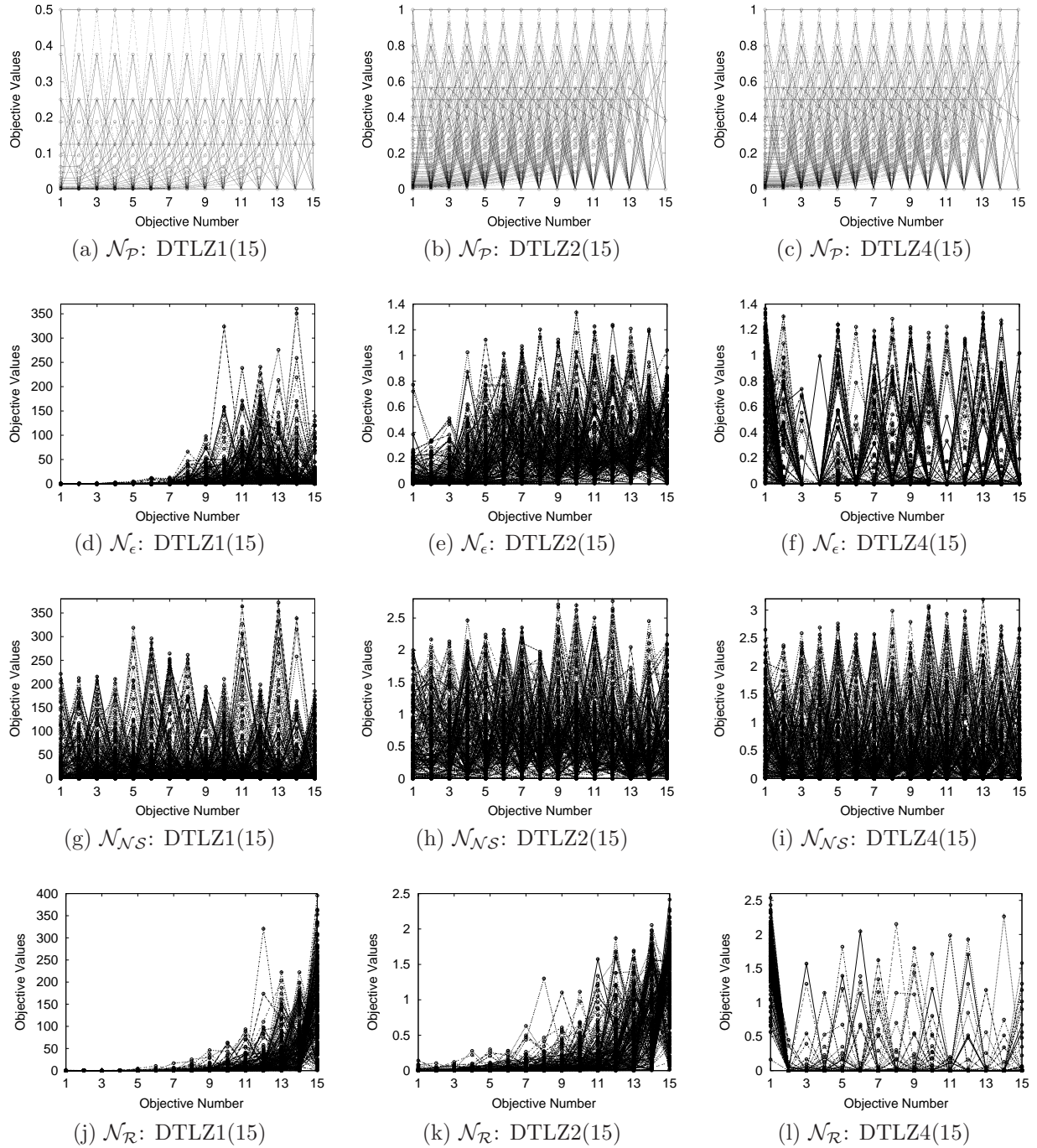


Figure 4.11: Non-redundant problems: On POF-approximation of  $\mathcal{N}_{\mathcal{P}}$ ,  $\mathcal{N}_{\epsilon}$ ,  $\mathcal{N}_{\mathcal{NS}}$  and  $\mathcal{N}_{\mathcal{R}}$  using parallel coordinate plots for DTLZ1(15), DTLZ2(15) and DTLZ4(15) problems.

The previous discussion highlights the differences in performance of the proposed objective reduction algorithms when applied to  $\mathcal{N}_\epsilon$ ,  $\mathcal{N}_{\mathcal{NS}}$  and  $\mathcal{N}_{\mathcal{R}}$ . An explanation of why the performance of L-PCA is superior to NL-MVU-PCA when applied to DTLZ1( $M$ ) and DTLZ3( $M$ ) problems follows. Consider DTLZ1(15) and the analysis is:

- The dataset in  $\mathcal{N}_\epsilon$  is mainly represented in a lower dimensional space, as shown in Figure 4.11d, due to the disparity between the objective values. Eventually, this disparity between the objective values is captured by the eigenvalues (as shown in Table 4.24b) leading to: (i) the first eigenvalue is very significant, (ii) the eigenvalues are unequal, and (iii) the majority of the variance is confined to a few eigenvalues. Nonetheless, it is expected from a non-redundant problem like DTLZ1 for all the eigenvalues to be equal. The previous argument is captured with  $\mathcal{N}_{\mathcal{P}}$  as shown in Table 4.24a, for both NL-MVU-PCA and L-PCA.
- The dataset in  $\mathcal{N}_\epsilon$  resembles a redundant problem as shown by Figure 4.11d and captured in Tables 4.24b and 4.24c, as contrary to  $\mathcal{N}_{\mathcal{P}}$ . During the objective reduction procedure, both NL-MVU-PCA and L-PCA have captured the correlated objective sets:  $\{f_1, \dots, f_5\}$ ,  $\{f_6, f_7\}$ , and  $\{f_8, f_9\}$ . However, owing to the redundancy in the dataset, NL-MVU-PCA was revealed to be more efficient in capturing it since: (i) the first eigenvalue is very significant  $e_1 = 0.44$  for NL-MVU-PCA ( $e_1 = 0.26$  for L-PCA); (ii) the eigenvalues are more unequal; and (iii) the number of significant eigenvalues is lower  $N_v = 8$  for NL-MVU-PCA ( $N_v = 15$  for L-PCA). Due to the captured eigenvalues the computed  $T_{cor}$  by NL-MVU-PCA is lower than the one obtained by L-PCA, as shown in Table 4.24c. This situation leads to  $m = 14$  for NL-MVU-PCA and  $m = M = 15$  for L-PCA.

It needs to be recognised that due to noised signal existent in  $\mathcal{N}_\epsilon$ , a non-redundant problem ( $m = M$ ), such as DTLZ1(15), can present characteristics that are associated with redundant problems ( $m < M$ ). In such cases, the objective reduction algorithms that are

Table 4.24: DTLZ1(15): Correlation-structure, eigenvalues and other highlights of L-PCA and NL-MVU-PCA corresponding to  $\mathcal{N}_{\mathcal{P}}$  and  $\mathcal{N}_{\epsilon}$

For brevity L-PCA and NL-MVU-PCA are abbreviated by “APR-A” and “APR-B”, respectively. For the same reason, Algorithm is abbreviated by “Alg.”

(a)  $\mathcal{N}_{\mathcal{P}}$ : Correlation matrix ( $R$ ), and eigenvalues of L-PCA and NL-MVU-PCA

Alg.	$e_1$	$e_2$	$e_3$	$e_4$	$e_5$	$e_6$	$e_7$	$e_8$	$e_9$	$e_{10}$	$e_{11}$	$e_{12}$	$e_{13}$	$e_{14}$	$e_{15}$
APR-A	0.08	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.00
APR-B	0.08	0.08	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.06	0.00

\* The correlation matrix satisfies Equation 4.10, in that, only the diagonal elements are positive, implying no two columns are identically correlated.

(b)  $\mathcal{N}_{\epsilon}$ : Correlation matrix ( $R$ ), and eigenvalues of L-PCA and NL-MVU-PCA

	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$f_7$	$f_8$	$f_9$	$f_{10}$	$f_{11}$	$f_{12}$	$f_{13}$	$f_{14}$	$f_{15}$
$f_1$	1.00	0.35	0.51	0.18	0.34	0.29	0.32	0.09	0.07	0.20	0.08	-0.01	-0.13	-0.08	-0.01
$f_2$	0.35	1.00	0.70	0.14	0.48	0.32	0.44	0.11	0.05	0.11	0.13	-0.01	-0.07	-0.08	-0.03
$f_3$	0.51	0.70	1.00	0.36	0.58	0.42	0.51	0.24	0.22	0.32	0.13	-0.09	-0.13	-0.10	-0.06
$f_4$	0.18	0.14	0.36	1.00	0.84	0.15	0.28	0.16	0.04	0.22	0.11	-0.06	-0.09	-0.04	-0.01
$f_5$	0.34	0.48	0.58	0.84	1.00	0.24	0.43	0.24	0.06	0.32	0.08	-0.07	-0.13	-0.08	-0.02
$f_6$	0.29	0.32	0.42	0.15	0.24	1.00	0.64	0.22	0.18	0.32	0.13	-0.04	-0.13	0.13	-0.02
$f_7$	0.32	0.44	0.51	0.28	0.43	0.64	1.00	0.34	0.24	0.33	0.09	-0.03	-0.13	0.04	-0.05
$f_8$	0.09	0.11	0.24	0.16	0.24	0.22	0.34	1.00	0.33	0.27	0.00	-0.14	-0.09	-0.06	-0.07
$f_9$	0.07	0.05	0.22	0.04	0.06	0.18	0.24	0.33	1.00	0.20	-0.06	-0.15	-0.06	-0.04	-0.12
$f_{10}$	0.20	0.11	0.32	0.22	0.32	0.32	0.33	0.27	0.20	1.00	0.08	-0.17	-0.17	-0.02	0.00
$f_{11}$	0.08	0.13	0.13	0.11	0.08	0.13	0.09	0.00	-0.06	0.08	1.00	-0.12	-0.09	0.01	-0.04
$f_{12}$	-0.01	-0.01	-0.09	-0.06	-0.07	-0.04	-0.03	-0.14	-0.15	-0.17	-0.12	1.00	-0.26	-0.04	-0.03
$f_{13}$	-0.13	-0.07	-0.13	-0.09	-0.13	-0.13	-0.13	-0.09	-0.06	-0.17	-0.09	-0.26	1.00	-0.06	-0.11
$f_{14}$	-0.08	-0.08	-0.10	-0.04	-0.08	0.13	0.04	-0.06	-0.04	-0.02	0.01	-0.04	-0.06	1.00	-0.02
$f_{15}$	-0.01	-0.03	-0.06	-0.01	-0.02	-0.02	-0.05	-0.07	-0.12	0.00	-0.04	-0.03	-0.11	-0.02	1.00
Alg.	$e_1$	$e_2$	$e_3$	$e_4$	$e_5$	$e_6$	$e_7$	$e_8$	$e_9$	$e_{10}$	$e_{11}$	$e_{12}$	$e_{13}$	$e_{14}$	$e_{15}$
APR-A	0.26	0.10	0.09	0.08	0.08	0.07	0.06	0.05	0.05	0.04	0.04	0.03	0.02	0.02	0.01
APR-B	0.44	0.25	0.13	0.07	0.05	0.03	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

(c)  $\mathcal{N}_{\epsilon}$ : Key highlights of L-PCA and NL-MVU-PCA

Alg.	$N_v$	$\mathcal{F}_e$	$\hat{\mathcal{S}}_i$	$T_{cor}$	$\mathcal{S}_i$	$\mathcal{F}_s$
APR-A	15	$\{f_1, \dots, f_{15}\}$	$\hat{\mathcal{S}}_1 = \hat{\mathcal{S}}_2 = \hat{\mathcal{S}}_3 = \hat{\mathcal{S}}_4 = \hat{\mathcal{S}}_5 = \{f_1, \dots, f_5\};$ $\hat{\mathcal{S}}_6 = \hat{\mathcal{S}}_7 = \{f_6, f_7\}; \hat{\mathcal{S}}_8 = \hat{\mathcal{S}}_9 = \{f_8, f_9\}$	0.94	$\emptyset$	$\{f_1, \dots, f_{15}\}$
APR-B	8	As above	As above	0.73	$\mathcal{S}_4 = \mathcal{S}_5 = \{f_4, f_5\}$	$\{f_1, \dots, f_4, f_6, \dots, f_{15}\}$

Table 4.25: Errors ( $\mathcal{E}_T$ ) associated with NL-MVU-PCA and L-PCA when applied to  $\mathcal{N}_\epsilon$ ,  $\mathcal{N}_{\mathcal{N}\mathcal{S}}$  and  $\mathcal{N}_{\mathcal{R}}$  for non-redundant test problems, with  $\theta = 0.997$

The mean ( $\mu$ ) and standard deviation ( $\sigma$ ) is obtained for 20 runs							
Test problems		NL-MVU-PCA			L-PCA		
DTLZ5( $I, M$ )		$\mathcal{N}_\epsilon$	$\mathcal{N}_{\mathcal{N}\mathcal{S}}$	$\mathcal{N}_{\mathcal{R}}$	$\mathcal{N}_\epsilon$	$\mathcal{N}_{\mathcal{N}\mathcal{S}}$	$\mathcal{N}_{\mathcal{R}}$
$I$	$M$	( $\mu \pm \sigma$ )	( $\mu \pm \sigma$ )	( $\mu \pm \sigma$ )	( $\mu \pm \sigma$ )	( $\mu \pm \sigma$ )	( $\mu \pm \sigma$ )
DTLZ1	05	0.00000 $\pm$ 0.00000	0.00000 $\pm$ 0.00000	0.000000 $\pm$ 0.000000	0.00000 $\pm$ 0.00000	0.00000 $\pm$ 0.00000	0.000000 $\pm$ 0.000000
	15	0.00058 $\pm$ 0.00070	0.00000 $\pm$ 0.00000	0.001430 $\pm$ 0.000922	0.00167 $\pm$ 0.00321	0.00000 $\pm$ 0.00000	0.000496 $\pm$ 0.001536
	25	0.00140 $\pm$ 0.00077	0.00000 $\pm$ 0.00000	0.001520 $\pm$ 0.000908	0.01728 $\pm$ 0.01241	0.00000 $\pm$ 0.00000	0.006605 $\pm$ 0.008947
DTLZ2	05	0.00000 $\pm$ 0.00000	0.00000 $\pm$ 0.00000	0.000000 $\pm$ 0.000000	0.00000 $\pm$ 0.00000	0.00000 $\pm$ 0.00000	0.000000 $\pm$ 0.000000
	15	0.00000 $\pm$ 0.00000	0.00000 $\pm$ 0.00000	0.000946 $\pm$ 0.000833	0.00000 $\pm$ 0.00000	0.00000 $\pm$ 0.00000	0.000177 $\pm$ 0.000772
	25	0.00002 $\pm$ 0.00007	0.00000 $\pm$ 0.00000	0.001803 $\pm$ 0.001432	0.00044 $\pm$ 0.00107	0.00000 $\pm$ 0.00000	0.005719 $\pm$ 0.005853
DTLZ3	05	0.00000 $\pm$ 0.00000	0.00000 $\pm$ 0.00000	0.000000 $\pm$ 0.000000	0.00000 $\pm$ 0.00000	0.00000 $\pm$ 0.00000	0.000000 $\pm$ 0.000000
	15	0.00022 $\pm$ 0.00033	0.00000 $\pm$ 0.00000	0.001286 $\pm$ 0.001403	0.00000 $\pm$ 0.00000	0.00000 $\pm$ 0.00000	0.000000 $\pm$ 0.000000
	25	0.00116 $\pm$ 0.00045	0.00000 $\pm$ 0.00000	0.001997 $\pm$ 0.000814	0.01167 $\pm$ 0.01102	0.00000 $\pm$ 0.00000	0.005978 $\pm$ 0.007417
DTLZ4	05	0.00000 $\pm$ 0.00000	0.00000 $\pm$ 0.00000	0.001997 $\pm$ 0.000814	0.00000 $\pm$ 0.00000	0.00000 $\pm$ 0.00000	0.000000 $\pm$ 0.000000
	15	0.00000 $\pm$ 0.00000	0.00000 $\pm$ 0.00000	0.000000 $\pm$ 0.000000	0.00000 $\pm$ 0.00000	0.00000 $\pm$ 0.00000	0.000000 $\pm$ 0.000000
	25	0.00000 $\pm$ 0.00000	0.00000 $\pm$ 0.00000	0.000000 $\pm$ 0.000000	0.00000 $\pm$ 0.00000	0.00000 $\pm$ 0.00000	0.000000 $\pm$ 0.000000

not efficient in capturing redundancy (like L-PCA and DRP-based greedy algorithm) might apparently perform better than more efficient algorithms (like NL-MVU-PCA and DRP-based exact algorithm).

As in the previous section the error reported in Table 4.25 needs to be interpreted in the light of: (a) the accuracy of the objective reduction algorithms; and (b) the quality of the POF-representation inherit in the non-dominated solution sets:

- The error obtained for  $\mathcal{N}_{\mathcal{N}\mathcal{S}}$  is lower as opposed to  $\mathcal{N}_\epsilon$ . This is attributed to a good POF-representation by  $\mathcal{N}_{\mathcal{N}\mathcal{S}}$ , while the POF-representation of  $\mathcal{N}_\epsilon$  is poor due to the disparity between the objective values, leading to  $m < M$ . Although the POF-approximation of  $\mathcal{N}_{\mathcal{N}\mathcal{S}}$  is considered to be poor, it does not imply that the POF-representation has to be poor as well, as argued in Section 4.8.4.
- The error obtained for  $\mathcal{N}_{\mathcal{R}}$  is for the majority of the cases higher than that obtained for  $\mathcal{N}_\epsilon$ . This can be attributed to a poor POF-representation by  $\mathcal{N}_{\mathcal{R}}$  when compared with the one obtained by  $\mathcal{N}_\epsilon$ . As an example consider DTLZ2(15) where: (i) for  $\mathcal{N}_{\mathcal{R}}$  there is a lack of conflict between most objectives as shown in Figure 4.11k, and (ii) for  $\mathcal{N}_\epsilon$  there is conflict between the majority of the objectives as shown in Figure 4.11e. This explains why more objectives have been omitted when the algorithms were applied to  $\mathcal{N}_{\mathcal{R}}$ .

- When applied to  $\mathcal{N}_\epsilon$ , NL-MVU-PCA shows less error than L-PCA. This can be attributed to more accurate principal components captured by NL-MVU-PCA due to the unfolding feature that is responsible for minimising the error incurred during an objective reduction.

## 4.11 Customisation of the Proposed Framework

The results shown in the previous sections have been generated using the most general parameters, such that, no *a priori* information regarding the nature of the problem is required. However, as mentioned in Section 4.2, the framework could be customised (with loss of generality) to handle either non-redundant ( $m = M$ ) or redundant problems ( $m < N$ ) in a more efficient way. The following two points present for each category of problems a way of customising the framework. Nonetheless, it needs to be recognised that this could be done in many other ways:

1. For highly redundant problems, the identification of conflicting objectives along each significant principal component ( $V_j$ ), during eigenvalues analysis (Section 4.1.2), is modified as follows:
  - (a) if objectives have different signs, the objectives with the most positive and most negative contribution to  $V_j$  are picked;
  - (b) if all objectives have the same sign then pick the objective with the highest magnitude contribution to  $V_j$  and also the second highest.
2. For non-redundant problems, it is recommended to skip the eigenvalues analysis and the remaining steps are kept unchanged. With this modification, the RCM analysis (Section 4.1.4) is performed directly in the original objective set ( $\mathcal{F}_0$ ).

In the discussion that follows the proposed customisations for each problem category are shown for DTLZ5(3,5) based on  $\mathcal{N}_\mathcal{P}$ :

Table 4.26: Customised framework for redundant problems: re-visiting DTLZ5(3, 5) corresponding to  $\mathcal{N}_{\mathcal{P}}$

(a) L-PCA: eigenvalues analysis (refer Table 4.7a)							
PCA	Variance	Cumulative	Objectives Selected				
$(N^v)$	(%)	(%)	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
1	68.67	68.67			$f_3$	$f_4$	
2	28.66	97.33				$f_4$	$f_5$
3	02.66	99.99				$f_4$	$f_5$
$\mathcal{F}_e = \{f_3, f_4, f_5\} \equiv \mathcal{F}_{\mathcal{T}}$							
(b) NL-MVU-PCA: eigenvalues analysis (refer Table 4.8a)							
PCA	Variance	Cumulative	Objectives Selected				
$(N^v)$	(%)	(%)	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
1	51.38	51.38				$f_4$	$f_5$
2	48.55	99.93			$f_3$		$f_5$
$\mathcal{F}_e = \{f_3, f_4, f_5\} \equiv \mathcal{F}_{\mathcal{T}}$							

1. When the problem is handled as redundant, the eigenvalues analysis is able to obtain an  $\mathcal{F}_{\mathcal{T}}$  for both L-PCA and NL-MVU-PCA, as shown in Table 4.26. Previously, in Tables 4.7a and 4.8a, no reduction was performed at this stage. As a result, with this modification the RCM analysis is no longer necessary.
2. When the problem is handled as non-redundant, the RCM analysis is able to reduce the objective set from  $\mathcal{F}_0$  to  $\mathcal{F}_{\mathcal{T}}$ , which makes the eigenvalues analysis look redundant.

The previous example shows how either eigenvalues analysis or RCM analysis can be used to obtain an  $\mathcal{F}_{\mathcal{T}}$ . However, even knowing that they might be applied independently, their simultaneous inclusion extends the applicability of the framework to a wider range of problems.

## 4.12 Parameter Sensitivity of the Proposed Framework

In this section the robustness and sensitivity of the proposed framework to different critical parameters is analysed. First, the analysis is on different  $\theta$ s (variance threshold). Reported

results are shown for  $\theta = 0.954$  and  $\theta = 0.682$  (selected in agreement with the three-sigma rule where 0.954 and 0.682 correspond to the probability of finding a solution within two and one standard deviations, respectively, from the mean in a Gaussian distribution), in Tables 4.27 and 4.28, respectively. Considering the discussion in Section 4.2 it is expected that a reduction in  $\theta$  would have the following effects:

1. For redundant problems: In this kind of problem only a few principal components need to be accounted for in order to capture most of the problem variance. Therefore, for an accurate algorithm capable of approximating the POF (such as NL-MVU-PCA) with good quality dataset (like  $\mathcal{N}_\epsilon$ ) the effect of reducing  $\theta$  in the performance should be minimal. In the case of inaccurate algorithms such as L-PCA and/or a bad quality set like  $\mathcal{N}_{\mathcal{S}}$  it is expected to reveal a higher deterioration of the results.
2. For non-redundant problems: In this case all the objectives are considered as essential hence any objective reduction is considered undesirable. In that sense, a lower  $\theta$  may prevent the eigenvalues analysis from picking all the objectives leading to inaccurate results. Therefore, the results deterioration for non-redundant problems will be more significant than with redundant problems.

The accuracy of the results follows the expected trend as shown in Tables 4.27 and 4.28 for  $\theta = 0.954$  and  $\theta = 0.682$ , respectively. The results show that even a significant change in  $\theta$  from 0.997 to 0.682 is not sufficient to alter the results. This further highlights the robustness of the framework.

Besides the variation in  $\theta$ , it is also important to investigate the robustness of the framework for different population sizes, since it is known that the number of solutions required to approximate the POF of a MaOP grows exponentially with  $M$ . As a result, the accuracy of the objective reduction algorithms can vary depending on the population size selected for analysis. DTLZ1 and DTLZ3 problems with 15 and 25 objectives are selected due to their poor performance in Table 4.22 for  $\mathcal{N}_\epsilon$ . The number of generations chosen for this analysis is



Table 4.27: Effect of the variance threshold ( $\theta = 0.954$ ) on the performance of the proposed L-PCA and NL-MVU-PCA algorithms on both redundant and non-redundant test problems, corresponding to  $\mathcal{N}_{\mathcal{N}\mathcal{S}}$  and  $\mathcal{N}_\epsilon$

The numbers in the table indicate the frequency of success in identifying the true  $I$  and  $\mathcal{F}_T$ , out of 20 runs. The dashes (–) replace 0 to imply inconsequential entries, as the prerequisite  $I$  is not met. The table’s footnote reports the problems that require multiple iterations of the algorithm to obtain accurate results, as: P—aR(bI), implying that for the problem P, a Runs out of 20, required b Iterations each.

DTLZ5( $I, M$ )		NL-MVU-PCA				L-PCA				$P(M)$	NL-MVU-PCA		L-PCA	
		$\mathcal{N}_\epsilon^a$		$\mathcal{N}_{\mathcal{N}S}^b$		$\mathcal{N}_\epsilon^c$		$\mathcal{N}_{\mathcal{N}S}$			$\mathcal{N}_\epsilon$	$\mathcal{N}_{\mathcal{N}S}$	$\mathcal{N}_\epsilon$	$\mathcal{N}_{\mathcal{N}S}$
		$I$	$\mathcal{F}_T$	$I$	$\mathcal{F}_T$	$I$	$\mathcal{F}_T$	$I$	$\mathcal{F}_T$					
2	05	20	20	20	20	20	04	20	01	1(05)	20	18	20	20
2	20	20	20	07	07	20	00	00	–	1(15)	06	20	15	20
2	50	20	20	14	14	20	00	09	00	1(25)	00	20	04	20
3	05	20	20	18	18	20	08	00	–	2(05)	20	20	19	20
3	20	20	20	00	–	20	00	00	–	2(15)	20	20	20	20
5	10	19	19	00	–	20	02	00	–	2(25)	16	20	16	20
5	20	20	20	00	–	19	00	00	–	3(05)	19	20	20	20
7	10	19	19	00	–	17	04	00	–	3(15)	11	20	18	20
7	20	16	16	00	–	16	03	00	–	3(25)	00	20	04	20
WFG3	05	20	20	20	20	20	19	20	03	4(05)	20	20	20	19
	15	15	15	20	20	09	00	19	00	4(15)	20	20	20	20
	25	09	09	20	20	04	01	19	00	4(25)	20	20	20	20

<sup>a</sup> DTLZ5(5, 10)—3R (2I); DTLZ5(5, 20)—6R (2I); DTLZ5(7, 10)—7R (2I); DTLZ5(7, 20)—8R (2I) and 1R (4I); WFG3(15)—6R (2I) and 3R (3I); WFG3(25)—5R (2I), 1R (3I) and 1R (4I).

<sup>b</sup> DTLZ5(2, 20)—1R (2I), 2R (4I), 2R (5I), 1R (6I), 1R (7I); DTLZ5(2, 50)—8R (2I) and 2R (3I); DTLZ5(3, 5)—9R (2I).

<sup>c</sup> DTLZ5(7, 20)—2R (2I); WFG3(5)—11R (2I); WFG3(15)—1R (3I).

Table 4.28: Effect of the variance threshold ( $\theta = 0.682$ ) on the performance of the proposed L-PCA and NL-MVU-PCA algorithms on both redundant and non-redundant test problems, corresponding to  $\mathcal{N}_{\mathcal{N}\mathcal{S}}$  and  $\mathcal{N}_\epsilon$

The numbers in the table indicate the frequency of success in identifying the true  $I$  and/or  $\mathcal{F}_T$ , out of 20 runs. The dashes (–) replace 0 to imply inconsequential entries, as the prerequisite  $I$  is not met. The table’s footnote reports the problems that require multiple iterations of the algorithm to obtain accurate results, as: P—aR(bI), implying that for the problem P, a Runs out of 20, required b Iterations each.

NL-MVU-PCA															L-PCA														
DTLZ5( $I, M$ )		$\mathcal{N}_\epsilon^a$				$\mathcal{N}_{NS}^b$				$\mathcal{N}_\epsilon^c$				$\mathcal{N}_{NS}^d$				$P(M)$	NL-MVU-PCA				L-PCA						
		$I$	$M$	$I$	$\mathcal{F}_T$	$I$	$\mathcal{F}_T$	$I$	$\mathcal{F}_T$	$I$	$\mathcal{F}_T$	$I$	$\mathcal{F}_T$	$N_\epsilon$	$N_{NS}$	$N_\epsilon$	$N_{NS}$												
2	05	20	20	20	20	20	01	20	18	1(05)	17	18	15	19															
2	20	20	20	08	08	20	00	02	01	1(15)	06	18	10	20															
2	50	20	20	19	19	20	00	20	02	1(25)	00	20	03	20															
3	05	20	20	18	18	20	05	01	01	2(05)	19	18	17	17															
3	20	20	20	00	–	20	00	00	–	2(15)	19	19	20	17															
5	10	17	17	00	–	18	03	00	–	2(25)	14	20	15	20															
5	20	19	19	00	–	10	00	00	–	3(05)	16	17	19	18															
7	10	16	15	00	–	18	07	00	–	3(15)	06	20	12	19															
7	20	13	13	00	–	09	01	00	–	3(25)	00	20	02	20															
WFG3	05	20	20	20	20	20	14	20	1	4(05)	15	20	19	18															
	15	15	15	20	20	03	01	20	0	4(15)	19	20	20	20															
	25	09	09	20	20	00	–	20	0	4(25)	20	20	20	20															

<sup>a</sup> DTLZ5(5, 10)—2R(2I); DTLZ5(5, 20)—6R(2I); DTLZ5(7, 10)—4R(2I); DTLZ5(7, 20)—5R(2I) and 1R(4I); WFG3(15)—6R(2I) and 3R(3I); WFG3(25)—5R(2I), 1R(3I) and 1R(4I).

<sup>b</sup> DTLZ5(2, 20)—1R(2I), 4R(3I), 1R(4I), 1R(5I) and 1R(7I); DTLZ5(2, 50)—11R(2I), 2R(3I) and 2R(4I); DTLZ5(3, 5)—10R(2I).

<sup>c</sup> DTLZ5(7, 10)—3R(2I); WFG3(5)—9R(2I); WFG3(15)—1R(2I).

<sup>d</sup> DTLZ5(2, 20)—1R(5I); DTLZ5(2, 50)—2R(2I).



Table 4.29: Effect of population size for  $\mathcal{N}_\epsilon$  with  $\theta = 0.997$  on the performance of NL-MVU-PCA

The results correspond to DTLZ1 and DTLZ3 test problems. The entries are formatted as  $\mu \pm \sigma$ , where  $\mu$  and  $\sigma$  represent the mean and standard deviation of the number of essential objectives identified, in 20 runs

Test problems	Population Size ( $N$ )				
	200	400	600	800	1000
DTLZ1(15)	06(14.1±0.7)	11(14.6±0.5)	14(14.7±0.5)	17(14.9±0.4)	18(14.9±0.3)
DTLZ1(25)	00(20.5±2.2)	02(21.0±2.2)	01(21.6±2.3)	03(22.8±1.9)	04(23.0±1.7)
DTLZ3(15)	11(14.3±0.9)	18(14.9±0.3)	17(14.9±0.4)	19(15.0±0.2)	20(15.0±0.0)
DTLZ3(25)	00(20.9±2.2)	00(22.3±1.9)	06(23.2±1.6)	01(22.8±1.5)	04(23.2±1.5)

given by  $N_g = 10N$ , where  $N$  is the population size. The extra computational time provided with an increase in the population size should provide an equivalent POF-approximation among the different runs. This implies that for  $N = \{200, 400, 600, 1000\}$  the corresponding number of generations is given by  $N_g = \{2000, 4000, 6000, 10000\}$ . The results are shown in Table 4.29 exhibit an higher improvement for  $M = 15$  and lesser improvement for  $M = 25$ . The results can be explained using Figure 4.12 where the solutions in  $\mathcal{N}_\epsilon$  are represented in parallel coordinate plots. In that, the disparity in the convergence levels between the objectives decreases with an increase in  $N$ . This highlights that the proposed algorithms' accuracy are as good as the data itself, in that, it is important to note that in case the population given for analysis is not representative of the problem characteristics, either by POF-representation or by POF-approximation, the accuracy of the algorithms cannot be guaranteed.

## 4.13 Real World Problems

In this section the proposed linear and nonlinear algorithms are applied to two real world problems and a comparison is made with DRP-based exact and greedy approaches.

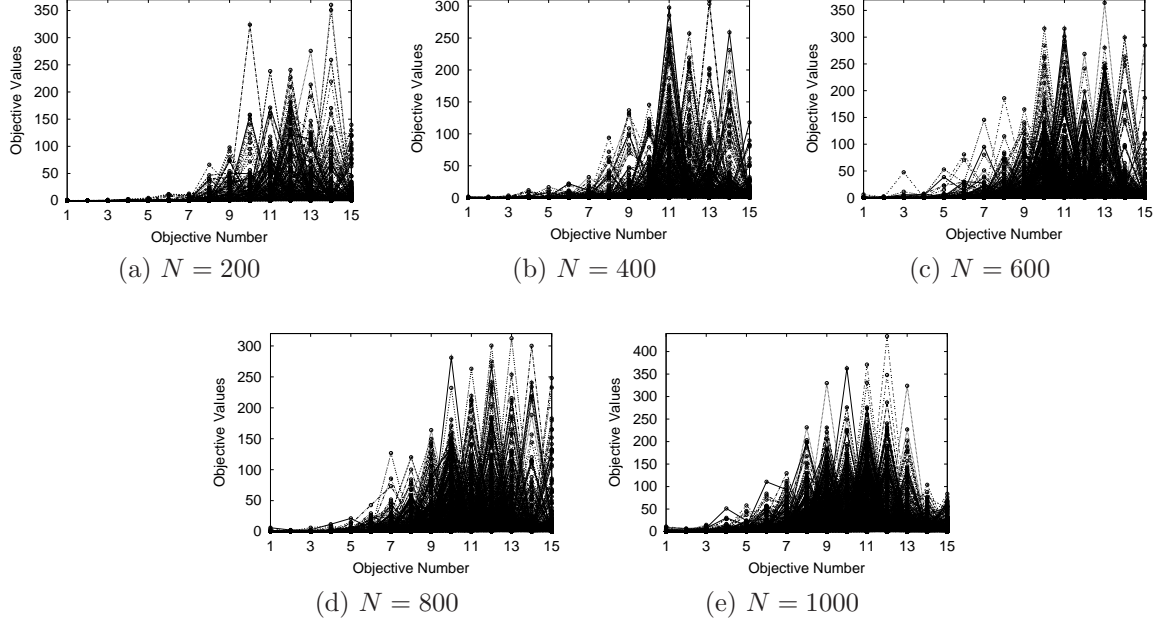


Figure 4.12: Effect of population size for  $\mathcal{N}_\epsilon$  with  $\theta = 0.997$  on the convergence level between the objectives depicted on parallel coordinate plots. The results correspond to DTLZ1(15) and a population size ranging from 200 to 1000 solutions.

#### 4.13.1 Multi-speed Gearbox Design Problem

This is a three-objective problem that does not belong to the many-objective category, however, it is used here to compare the proposed algorithms vis-à-vis with the DRP-based approaches. The aim of the problem is to optimise the design of a multi-speed gearbox. While the objectives and constraints formulations can be found in [159], it may be noted that the problem comprises of five linear equality constraints, 96 nonlinear inequality constraints and 24 variables. Due to the problem's complexity, the variables are divided in three different categories as such: (i) 10 real variables, (ii) 18 integer variables and (iii) one discrete variable. Also to be noted is the physical meaning of the objectives as follows:

$$\left. \begin{aligned}
 &\text{Minimise } f_1(\mathbf{x}) \equiv \text{overall volume of gear material used,} \\
 &\quad \quad \quad \text{(which is directly related to the weight and cost of the gearbox),} \\
 &\text{Maximise } f_2(\mathbf{x}) \equiv \text{power delivered by the gearbox,} \\
 &\text{Minimise } f_3(\mathbf{x}) \equiv \text{the centre distance between input and output shafts.}
 \end{aligned} \right\} \quad (4.13)$$

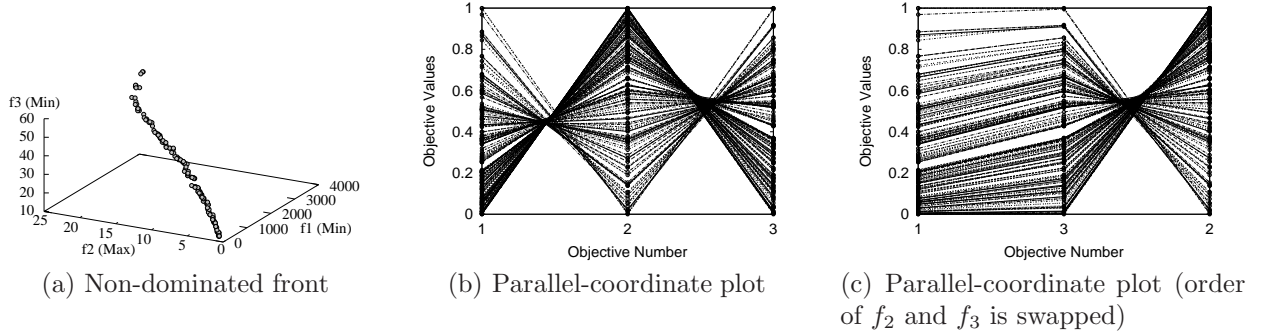


Figure 4.13: Multi-speed gearbox design problem: The plots correspond to one run of NSGA-II.

Table 4.30: Performance of objective reduction algorithms on two real-world problems

These results correspond to 20 NSGA-II runs with uniformly distributed seeds; each run corresponding to 200 population size and 2000 generations

Real-world Problems	Proposed approaches		DRP [93, 2]: $\delta$ -MOSS, 0% Error	
	NL-MVU-PCA	L-PCA	Greedy	Exact
(a) Multi-speed gearbox <sup>a</sup>	$\{f_1, f_2\}$	$\{f_1, f_2\}$	$\{f_1, f_2, f_3\}$	$\{f_1, f_2, f_3\}$
(b) Storm drainage system <sup>b</sup>	$\{f_2, f_3, f_4, f_5\}$	$\{f_1, f_2, f_4, f_5\}$ <sup>c</sup>	$\{f_2, f_3, f_4, f_5\}$	$\{f_2, f_3, f_4, f_5\}$

<sup>a</sup> The error associated with NL-MVU-PCA and L-PCA is  $0.00262 \pm 0.00035$  and  $0.00559 \pm 0.00074$ , respectively.

<sup>b</sup> The error associated with NL-MVU-PCA and L-PCA is  $0.00002 \pm 0.00001$  and  $0.00021 \pm 0.00012$ , respectively.

<sup>c</sup> In 18 out of the 20 runs, L-PCA finds  $\mathcal{F}_s = \{f_1, f_2, f_4, f_5\}$ , while twice it finds  $\mathcal{F}_s = \{f_2, f_3, f_4, f_5\}$ .

In Figure 4.13a the Pareto-front approximation obtained by  $\mathcal{N}_{NS}$  with 200 population size and a single run for 2000 generations is generated for the multi-speed gearbox design problem. While the conflict between the objectives is not evident in Figure 4.13a the parallel coordinate plot in Figure 4.13c shows that objectives  $f_1$  and  $f_3$  are non-conflicting since it cannot lead to incomparable solutions. The obtained results can be physically justifiable because, for a fixed number of gears, the lower the centre distance between the input and output shafts (i.e.  $f_3$ ), the smaller the size of each gear will be, resulting in lower overall volume of gear material (i.e.  $f_1$ ).

The objective reduction results obtained by the algorithms is reported in Table 4.30. While L-PCA and NL-MVU-PCA have identified objective  $f_3$  as redundant leading to reduced set  $\{f_1, f_2\}$  both DRP algorithms, greedy and exact, have failed to capture it and

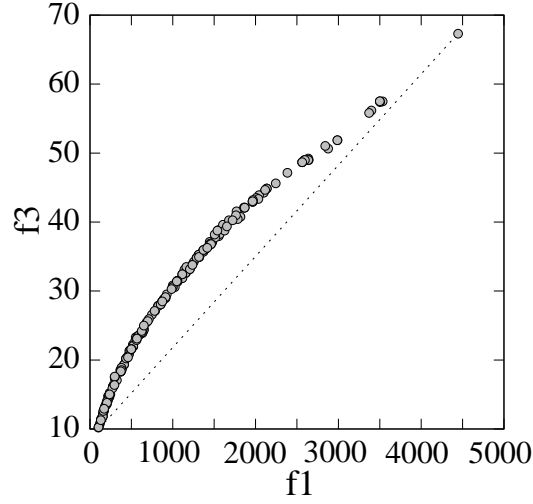


Figure 4.14: Highlighting the nonlinear and linear characteristics of gearbox design problem. Nonlinearity in  $f_1$ - $f_3$  subspace. One NSGA-II run.

instead the whole objective set is considered in conflict.

The correlation between objectives  $f_1$  and  $f_3$  is better captured in Figure 4.14 where a growth in objective  $f_1$  means also a growth objective  $f_3$ . While L-PCA and NL-MVU-PCA algorithms are able to exploit the positive correlation between objective  $f_1$  and  $f_3$  the DRP algorithms fail in doing so. The failure of DRP is attributed to only three solutions out of 20 (Figure 4.13c) that show conflict between  $f_1$  and  $f_3$  that have been filtered by NL-MVU-PCA as noise.

### 4.13.2 Storm Drainage System Problem

The storm drainage system problem is a five-objective problem with seven constraints that relates to the optimisation of a storm drainage system in an urban area. Even knowing that a detailed description of the problem can be found in [160] the physical meaning of the

objectives and constraints is captured by Equation 4.14.

$$\begin{array}{ll}
 \text{Minimise}_{(x_1, x_2, x_3)} & \left. \begin{array}{l}
 f_1(\mathbf{x}) \equiv \text{Drainage network cost,} \\
 f_2(\mathbf{x}) \equiv \text{Storage facility cost,} \\
 f_3(\mathbf{x}) \equiv \text{Treatment facility cost,} \\
 f_4(\mathbf{x}) \equiv \text{Expected flood damage cost,} \\
 f_5(\mathbf{x}) \equiv \text{Expected economic loss due to flood,} \\
 \\
 \text{subject to:} \\
 g_1(\mathbf{x}) \equiv \text{Average no. of floods/year,} \\
 g_2(\mathbf{x}) \equiv \text{Probability of flood depth exceeding 1 basin-inch,} \\
 g_3(\mathbf{x}) \equiv \text{Average no. of pounds/year of suspended solids,} \\
 g_4(\mathbf{x}) \equiv \text{Average no. of pounds/year of settleable solids,} \\
 g_5(\mathbf{x}) \equiv \text{Average no. of pounds/year of } BOD, \\
 g_6(\mathbf{x}) \equiv \text{Average no. of pounds/year of } N, \\
 g_7(\mathbf{x}) \equiv \text{Average no. of pounds/year of } P_{04}.
 \end{array} \right\} \quad (4.14)
 \end{array}$$

For this problem the results reported in Table 4.30 shows that either  $f_1$  or  $f_3$  is considered redundant by the algorithms. These results are correct due to the following analysis:

- the objectives  $f_1$  and  $f_3$  are positively correlated as evident in Figure 4.15b since it cannot lead to incomparable solutions,
- and the  $\mathcal{N}_{\mathcal{NS}}$  obtained as shown in Figure 4.15c with objective set  $\{f_2, f_3, f_4, f_5\}$  conforms with the original  $\mathcal{N}_{\mathcal{NS}}$  in Figure 4.15a.

In this problem the DRP algorithms have identified the correct problem dimensionality due to the linearity of the problem as evident in Figure 4.16. Furthermore, NL-MVU-PCA selected  $f_3$  as an essential objective since the variance of  $f_3$  is higher than the one from  $f_1$ . It is possible to observe this in Figure 4.16. However, the same could only be captured by L-PCA in two cases out of 20 runs.

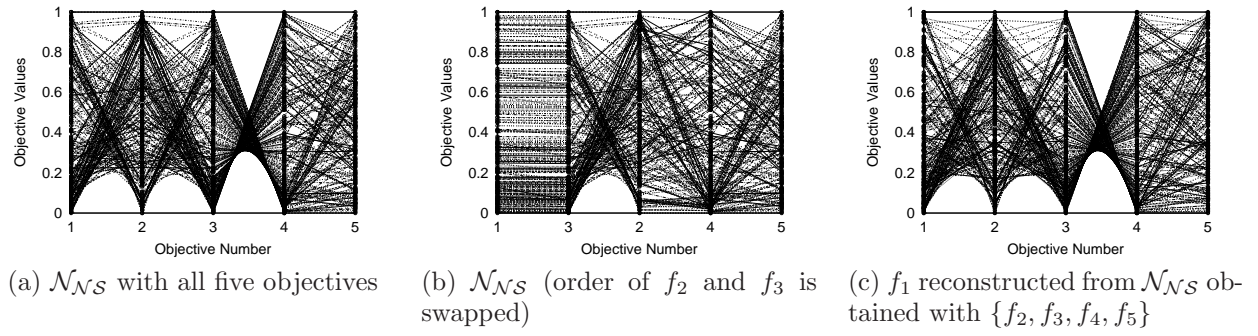


Figure 4.15: Storm drainage system problem: Parallel coordinate plots (normalised), corresponding to one run of NSGA-II.

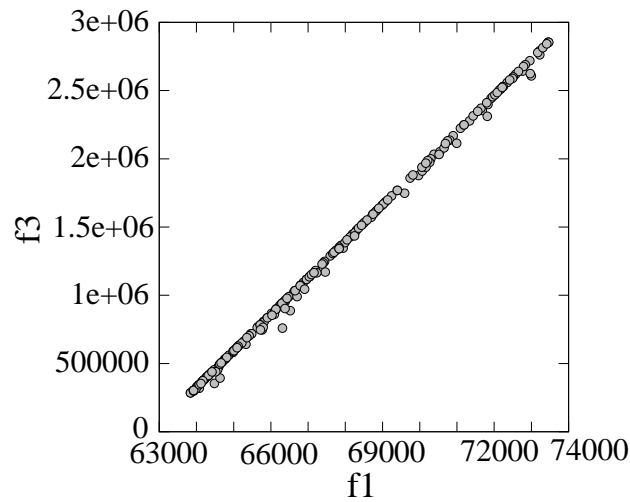


Figure 4.16: Highlighting the nonlinear and linear characteristics storm drainage system problem. Linearity in  $f_1$ – $f_3$  subspace. One NSGA-II run.

## 4.14 Comparative Analysis of the Proposed and DRP-based Objective Reduction Algorithms

The experimental results, shown in the previous sections, have highlighted that the proposed and the DRP-based objective reduction algorithms are able to identify accurately the essential objective set ( $\mathcal{F}_{\mathcal{T}}$ ) for a wide range of redundant and nonredundant test problems when applied to solution sets characterised by exact-optimal solutions, i.e., no noise ( $\mathcal{N}_{\mathcal{P}}$ ).

A solution set generated by an MOEA is characterised by exact-optimal (intrinsic dimensionality  $m = |\mathcal{F}_{\mathcal{T}}|$ ) and non-exact optimal solutions (dimensionality different from  $m$ ). It is therefore important for objective reduction algorithms to be capable of handling noisy solution sets that are generated by MOEAs and thus to be able to identify the essential objective sets for a range of problems, with accuracy, and computational efficiency.

### 4.14.1 Strength and Limitations of the Different Algorithms

The experimental results that correspond to noisy solution sets (namely  $\mathcal{N}_{\mathcal{NS}}$  and  $\mathcal{N}_{\epsilon}$ ), reveal the following:

1. DRP-based objective reduction algorithms have failed to identify an essential objective set for redundant and non-redundant problems. The algorithms have identified a higher and lower dimension than  $m$  for redundant and nonredundant problems, respectively. This reveals that DRP-based objective reduction algorithms are very sensitive to noise which can be attributed to the dominance relation preservation. As a result, a solution set with a single noisy solution prevents these algorithms to capture the correct problem dimensionality when  $\delta = 0$  is requested. As an example, consider the multi-speed gearbox design problem (Section 4.13.1), where three noisy solutions out of 200 have prevented the algorithms from identifying a redundant objective. Instead it was considered non-redundant.

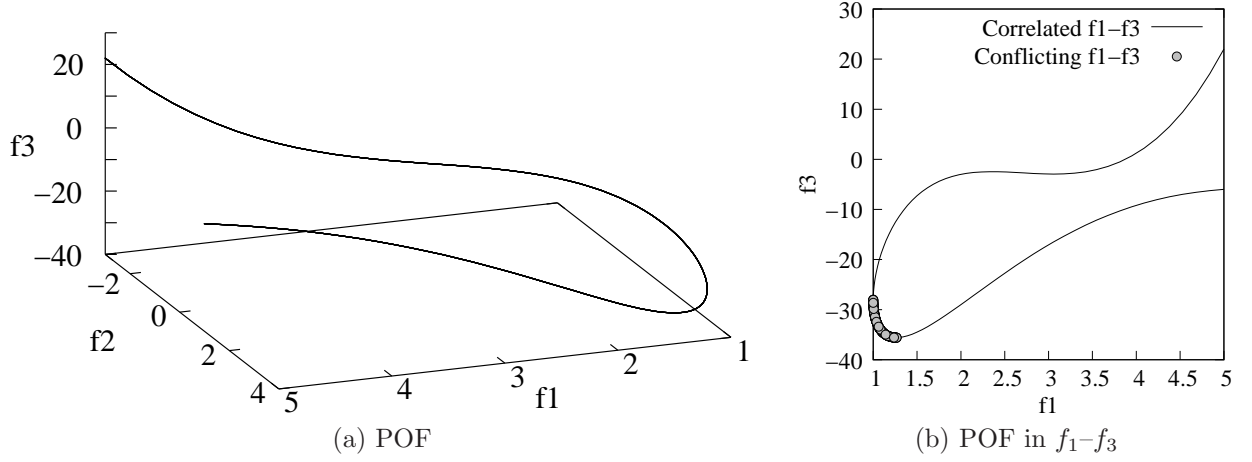


Figure 4.17: Splines problem:  $f_1 = x^2 + 1$ ,  $f_2 = -x^2 + x + 3$ , and  $f_3 = -(f_1 + f_2^3)$ , where  $x \in [-2, 2]$ . The analysis is based on  $\mathcal{N}_{\mathcal{P}}$ .

2. The L-PCA and NL-MVU-PCA, are based on preserving the correlation-structure instead of the dominance relations of the solution set. To handle noisy solution sets, the algorithms account for 99.7% of the variance ( $\theta = 0.997$ ). They only consider those correlations between objectives, with a strength, equal or superior to a problem-specific dynamically generated eigenvalue-based threshold ( $T_{cor}$ ). The *de-noising* strategy adopted by these algorithms allows them to identify an essential objective set, with reasonably accuracy for a range of problems.

The advantages of preserving the correlation-structure when dealing with noisy solution sets are clear for the previous problems due to the *de-noising* effect. However, in some situations, the *unnoise* operation might exclude some important problem features that were mistakenly thought to be noise. Consider a situation where two objectives can be globally correlated and some conflict exists only locally. When the existent conflict is not significant enough, it might be interpreted as noise. As an example, consider a three-objective problem whose conflict and correlations iterations between  $f_1$  and  $f_3$  are represented in Figure 4.17.

For this problem the analysis of L-PCA and NL-MVU-PCA is represented in Tables 4.31 and 4.32, corresponding to the correlation matrix ( $R$ ) and key highlights, respectively. During the RCM analysis (Section 4.1.4) the objectives  $f_1$  and  $f_3$  have been considered *potentially*



Table 4.31: Splines problem: Correlation matrix ( $R$ ) for  $\mathcal{N}_{\mathcal{P}}$

	$f_1$	$f_2$	$f_3$
$f_1$	1.0000	-0.7201	0.6942
$f_2$	-0.7201	1.0000	-0.8990
$f_3$	0.6942	-0.8990	1.0000

Table 4.32: Splines problem: Key highlights of L-PCA and NL-MVU-PCA for  $\mathcal{N}_{\mathcal{P}}$

Features	L-PCA	NL-MVU-PCA
$e_i$	0.848; 0.117; 0.033	0.993; 0.006; 0.000
$sc_i$	0.561; 0.570; 0.574	0.305; 0.506; 0.803
$N_v$ and $M_{2\sigma}$	3 and 2	3 and 1
$\mathcal{F}_\epsilon$	$\{f_1, f_2, f_3\}$	$\{f_1, f_2, f_3\}$
$\hat{\mathcal{S}}_i$	$\hat{\mathcal{S}}_1 = \hat{\mathcal{S}}_3 = \{f_1, f_3\}$ , where $R_{13} = 0.6942$	
$T_{cor}$	0.7171	0.3375
$\mathcal{S}_i$	$\mathcal{S}_1 = \mathcal{S}_3 = \emptyset$	$\mathcal{S}_1 = \mathcal{S}_3 = \{f_1, f_3\}$
$\mathcal{F}_s$	$\{f_1, f_2, f_3\}$	$\{f_2, f_3\}$
Error: $\mathcal{E}_0$	0.0	2.89%

*identically* correlated (Equation 4.1(i)) and further confirmation depends on whether the correlation strength is higher or lower than  $T_{cor}$  (Equation 4.1(ii)). Due to the unfolding by NL-MVU-PCA the variance of the problem is mostly captured by the first eigenvalue, when compared to L-PCA, which results in a lower  $T_{cor}$ . The subsequent analysis leads to:

- for L-PCA since  $R_{13} < T_{cor}$ , then  $f_1$  and  $f_3$  are considered uncorrelated which prevents any of them to be removed, leading to  $\mathcal{F}_s = \{f_1, f_2, f_3\}$ ,
- for NL-MVU-PCA since  $R_{13} > T_{cor}$ , then  $f_1$  and  $f_3$  are considered correlated which results in the elimination of  $f_1$ , leading to  $\mathcal{F}_s = \{f_2, f_3\}$  corresponding to an error of 2.89%.

Objective  $f_1$  is responsible for a part of the POF as represented in Figure 4.17a, therefore its elimination is undesirable. Like L-PCA also the DRP-based algorithms have identified  $\mathcal{F}_s = \{f_1, f_2, f_3\}$ . However, the *de-noising* capabilities of NL-MVU-PCA may over-reduce the objectives if some problem features may resemble noise, which here leads to  $\mathcal{F}_s = \{f_2, f_3\}$ . This limitation could be countered by clustering the objective space in different clusters and then by applying these algorithms to each cluster. However, that could require additional

computational cost.

#### 4.14.2 Computational Complexity of Objective Reduction Algorithms

In this section the computational complexity of different objective reduction algorithms is analysed. Let  $M$  denote the number of objectives and  $N$  is the size of the non-dominated set. For the proposed algorithms the computational complexity is as follows:

- The computational complexity of L-PCA is dictated by: (i) the correlation matrix ( $R$ ) computation, and (ii) the eigendecomposition of the same matrix. The number of objective pairs involved in the correlation matrix computation is  $M(M - 1)/2$ , which leads to  $M^2$  for a single solution and  $NM^2$  for  $N$  solutions. For the eigendecomposition, an approximate for a  $M \times M$  symmetric matrix is known to be  $O(M^3)$ . As a result, the computational complexity of L-PCA is  $O(NM^2 + M^3)$ .
- For NL-MVU-PCA it is necessary to ascertain the kernel matrix ( $K$ ) by posing it as a semidefinite programming problem. The computational complexity of semidefinite programming is reported in [5] to be  $O(n^3 + c^3)$ , where  $n$  is the matrix size and  $c$  is the number of constraints. Here, the matrix is  $M \times M$  and the number of constraints necessary to maintain the local *isometry* is  $Mq$ . This leads to  $O(M^3 + M^3q^3)$  which collapses to  $O(M^3q^3)$ . In the worst case, where  $q = M - 1$ , the computational complexity is  $O(M^6)$ .

The Dominance Relation Preservation (DRP) approach computational complexity is the most demanding since the exact approach complexity increases exponentially in  $M$  and is quadratic in  $N$ , which leads to  $O(N^2M2^M)$ . Also the greedy approach is likely to be more expensive than NL-MVU-PCA (worst case is  $O(M^6)$ , for  $q = M - 1$ ), which is reported to be  $O(\min\{N^2M^3, N^4M^2\})$  by the authors in [2]. The Unsupervised Feature Selection (UFS) approach like L-PCA requires the computation of a correlation matrix which leads to

Table 4.33: Computational complexity of objective reduction algorithms

$M$ denotes the number of objectives and $N$ the size of the non-dominated set	
Approaches	Computational complexity
A. Dominance relation preservation ( $\delta$ -MOSS)	
(i) Exact Algorithm	$O(N^2 M 2^M)$
(ii) Greedy Algorithm	$O(\min\{N^2 M^3, N^4 M^2\})$
B. Unsupervised feature selection	$O(NM^2) + \text{clustering overhead}$
C. Removal of data dependencies	
(i) PCA based reduction	$O(NM^2 + M^3)$
(ii) MVU-PCA based reduction	$O(M^3 q^3)$ where $q$ is the <i>neighbourhood size</i> <sup>a</sup>

<sup>a</sup> In the most constrained case,  $q = O(M)$ , which leads to  $O(M^6)$ .

$O(NM^2)$ . Also, it is necessary to divide the population in neighbourhoods of a determined size around each objective which can be achieved by a clustering algorithm. As a result the UFS computational complexity is  $O(NM^2) + \text{clustering overhead}$ . The previous analysis is summarised in Table 4.33.

### 4.14.3 Key Inferences

Based on the results shown in previous sections the inferences drawn from the application of objective reduction to MaOPs are as follows:

1. To identify a smallest set of  $m$  ( $m \leq M$ ) conflicting objectives, the DRP-based algorithms preserve the dominance relations of the solution set, which is itself questionable in the presence of solution sets with noise. As a consequence, the application of DRP to solution sets generated by MOEAs is inappropriate, since they are characterised by noise. This addressed one research gap identified in Section 2.4.1, since it can be inferred that preservation of the correlation-structure is a better criterion than preservation of the dominance relations when dealing with solutions sets with noise.
2. The usage of DRP-based algorithms in finding a  $\delta$ -minimum objective set for  $\delta = 0$  is impractical for problems with large  $M$ . The exact algorithm which guarantees a 0-minimum objective set suffers from high computational complexity, and the greedy

algorithm does not guarantee a 0-minimum objective set.

3. The proposed objective reduction algorithms, namely L-PCA and NL-MVU-PCA, preserve the correlation-structure of the solution set as against the dominance relations. The preservation of the correlation-structure together with the de-noising capabilities makes the proposed objective reduction algorithms capable of handling solution sets generated by MOEAs, that is to say, characterised by noise. Furthermore, in case the correlations between objectives that characterised a good POF-representation are not fully captured, the algorithms may still identify an  $\mathcal{F}_T$  due to the iterative approach. Also, if some objectives that are conflicting in the POF but are identified as correlated might be still considered as conflicting due to the de-noising features based on  $T_{cor}$ , which leads to more accurate results. This addresses the research gap identified in Section 2.4.1, since the *de-noising* strategy adopted by the proposed algorithms is capable to adapt itself based on the problem in consideration and to handle solution sets generated by MOEAs.
4. A more promising manner to handle problems with  $M \gg 4$ , could be to: (i) employ the proposed algorithms to identify an essential objective set from a solution set (generally noisy if generated by an MOEA), then (ii) apply the DRP-based algorithms to the filtered solutions selected by the *de-noising* feature, which are considered close to the exact-optimal solutions. With this approach it is possible to inform the decision maker of the different objective subsets that correspond to different requested  $\delta$ -errors in their dominance structure.

## 4.15 Summary

In this chapter a machine learning based framework for linear and nonlinear objective reduction algorithms, namely L-PCA and NL-MVU-PCA, has been proposed. The proposed objective reduction algorithms (also PCA-based algorithms) are capable of finding an essen-

tial objective set by preserving the correlation–structure of the given non-dominated solution set and to assess the quality of the results using an error measure. The proposed framework is based on previously published works [6] and [7], and the distinctive contribution relates to:

1. The four goals that the framework pursues:
  - (a) Generality: the scope in [6] and [7] is limited to highly redundant problems, where  $m \ll M$ . The scope of the new proposed algorithms is extended to include: (i) problems with moderate redundancy, where  $m < M$ ; and (ii) problems with low or negligible redundancy, where  $m \approx M$ . This allows the framework to handle cases with different degrees of redundancy.
  - (b) De-noising the input solution set: the proposed framework acknowledges that the input data may be characterised by *noise*. *Noise* in this context is defined as the difference between the dominance relations of the solutions on the POF and that obtained by the given solutions. The *de-noising* strategy (inexistent in [6] and [7]) employs an eigenvalue based problem-specific correlation threshold  $T_{cor}$ , and two pairs of objectives are considered as correlated only if their strength of correlation exceeds  $T_{cor}$ .
  - (c) Addition: the additions over [6] and [7] are: (i) the correlation threshold  $T_{cor}$  that serves the *de-noising* strategy as mentioned above; (ii) the objective contribution score  $sc_i$ , utilised to rank the objectives accordingly to their contribution to the significant principal components, during the RCM analysis; and (iii) proposition of an error measure that allows an assessment of the quality of the results obtained.
  - (d) Parameter reduction: in comparison with [6] and [7] the number of parameters have been reduced which includes: (i) the threshold cut (TC) has been removed during eigenvalues analysis; and (ii) during the unfolding process performed by maximum variance unfolding it is recommended to set  $q = M - 1$  for all purposes

since it gives sufficient degree of freedom while preserving the local isometry.

2. Extensive simulations and results: the presented results correspond to over 9000 simulations performed on 24 versions of six test problems and two real-world problems. The simulations are conducted for 20 runs each and include solution sets generated: (i) on the true POF, namely  $\mathcal{N}_{\mathcal{P}}$ ; (ii) by NSGA-II, namely  $\mathcal{N}_{\mathcal{NS}}$ ; and (iii) by  $\epsilon$ -MOEA, namely  $\mathcal{N}_{\epsilon}$ . The simulations conducted for the latter two solution sets are repeated for three different parameter settings of the proposed framework, and also for two algorithms based on dominance relation preservation (DRP). This extensive setup has allowed for the proposed algorithms to be evaluated in terms of their critical parameters and on the quality of POF-approximation (convergence and diversity) and POF-representation (correlation-structure) provided by the underlying non-dominated solution set.
3. Performance comparison: the proposed algorithms are compared against an alternative approach based on dominance relation preservation (DRP). A comparative analysis between the two approaches focuses on their strengths, limitations and corresponding computational complexity.

Based on the results obtained in this chapter, the following can be said:

- The PCA-based and the DRP-based algorithms have accurately identified an essential objective set for redundant and non-redundant problems when applied to solution sets on the true POF characterised by no noise (corresponding to  $\mathcal{N}_{\mathcal{P}}$ ). The only exception is L-PCA when applied to redundant problems since it could not identify the particular essential objective set that accounts for the highest variance. This lack of accuracy can be attributed to the limitations that PCA suffers from, that were previously highlighted in Section 2.3.2.
- The PCA-based algorithms are capable of handling non-dominated sets with noise (corresponding to  $\mathcal{N}_{\mathcal{NS}}$  and  $\mathcal{N}_{\epsilon}$ ) better than the DRP-based algorithms, for redundant and

non-redundant problems. The latter approach could not identify an essential objective set for redundant and non-redundant problems. This reveals that objective reduction algorithms based on DRP are very sensitive to noise which can be attributed to the criterion of preserving the dominance relations. In that, it is inferred that a solution set with a single noisy solution is able to directly influence the global dominance relation of the entire solution set. On the other hand, the PCA-based algorithms preserve the correlation-structure as opposed to the dominance relations of the solution set. For handling the noisy solution sets, the PCA-based algorithms proceed as follows: (i) account for 99.7% of the variance, and (ii) only consider those correlations between objectives, with a strength, equal or superior to a problem-specific dynamically generated eigenvalue-based threshold ( $T_{cor}$ ). The adopted *de-noising* strategy allows for an essential objective set to be identified, with reasonable accuracy for a wide range of problems with different degrees of redundancy.

Moreover, it is found that if the correlation-structure of the solution set conforms with the correlation-structure of the POF, accurate results by the PCA-based algorithms are expected. A solution set with such a feature is identified as having a good POF-representation. Based on this observation it is established that: (i) a good POF-approximation is a sufficient, but not a necessary condition for a good POF-representation and (ii) a good POF-representation is not necessarily a good POF-approximation. Hence, the assumption that a good POF-representation signifies accurate results could be captured. This explains why the PCA-based algorithms when applied to some non-dominated sets with poor POF-approximation could identify correctly the problem essential objective set.

## Chapter 5

# Decision Support Based on *Offline* Objective Reduction Framework

In the previous chapter, a machine learning based framework for linear and nonlinear objective reduction algorithms, namely L-PCA and NL-MVU-PCA, has been proposed. The approach is able to identify a smallest set of  $m$  ( $m \leq M$ ) conflicting objectives which generates the same POF as the original problem by preserving the correlation-structure of the original problem. When compared against an alternative approach based on preserving the dominance relations of the original objective set, namely dominance relation preservation (DRP), the proposed algorithms revealed higher accuracy and robustness when applied to problems with different degrees of redundancy.

In this chapter the concepts of  $\delta$ -MOSS and  $k$ -EMOSS, defined as the smallest objective sets corresponding to pre-specified  $\delta$  errors and the objective sets of pre-specified size  $k$  that corresponds to minimum error, respectively, are explored. Towards this, a machine learning based framework to introduce *objectivity*, *repeatability*, *consistency* and *coherence* in the DM's preferences is proposed. The proposed framework is an extension of Framework 1 from Chapter 4, since it also reveals the smallest set of conflicting objectives which can generate the same POF as that by the original problem. Besides that, the framework new features



includes: revelation of an essential objective set, preference-ranking of all the objectives,  $\delta$ -MOSS ( $\delta$ -Minimum Objective Subset) and  $k$ -EMOSS (Minimum Objective Subset of Size  $k$  with Minimum Error) analysis, and a simple yet meaningful visual representation of all the previous analysis.

In this chapter the focus is on demonstrating how the propose framework can be utilised as a decision support for the DM, and beside that, its performance is compared vis-à-vis with an alternative approach based on preserving the dominance relation, namely dominance relation preservation (DRP). It will be revealed that the obtained results by the decision support based on the proposed framework are more reliable than the ones obtained by DRP. In that, the lack of accuracy of the DRP based algorithms could perhaps be explained by the limitations that the method suffers from.

## 5.1 Rationale for a Decision Support to Deal with Many-objective Optimisation Problems

Given the difficulties associated with human decision making identified in Section 2.2, there is a necessity of a decision support for DMs to articulate their preferences with rationality when dealing with MaOPs. To achieve this purpose, the proposed decision support is based on the premise that a logical implementation can be achieved through learning the preference-structure of the objective functions, inherent in the formulation of the optimisation problem. The justification for this premise is given below.

Generally in multi-objective optimisation an optimisation task is compromised of three phases (Figure 5.1): (i) modelling, (ii) search/optimisation, and (iii) decision-making. One of the main features of MOEAs is that no *a priori* information needs to be given by the DM that would influence the optimisation/search ability of the MOEA. Instead, the decision maker phase (Phase-3) only takes place after the optimisation phase (Phase-2). As opposed to the previous sequence, the MCDM based approaches allow the use of the DM's preferences

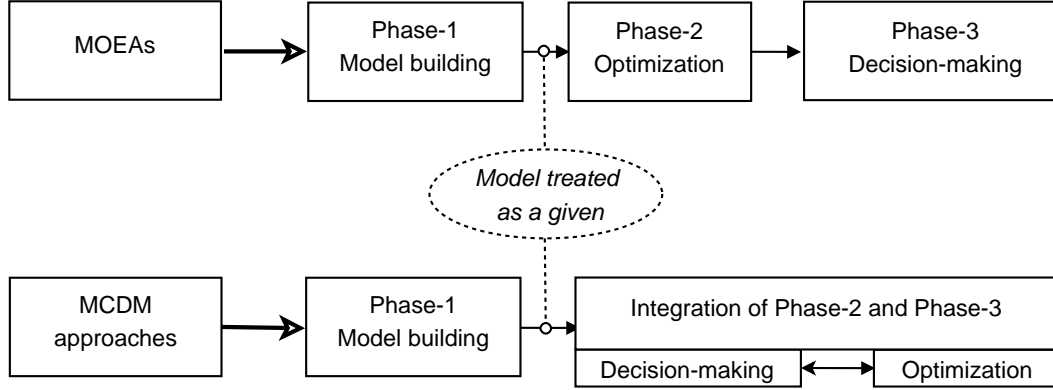


Figure 5.1: On the difference between the MOEA and MCDM approaches.

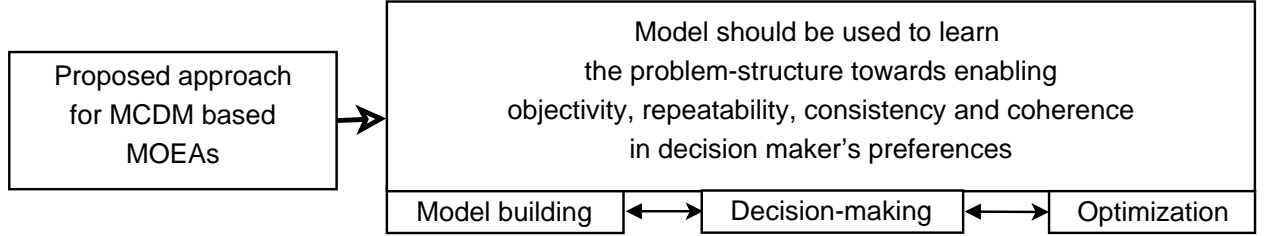


Figure 5.2: Proposed approach on introducing *objectivity, repeatability, consistency* and *coherence* in DM's preferences, in the context of MCDM based MOEAs.

before the optimisation/search (*a priori* approach) or intermittently during the optimisation/search [161] (*interactive* or *progressive* approach). This approach can also be viewed as a form of integration between the optimisation phase (Phase-2) and the decision-making phase (Phase-3).

In any of the models described above, the model phase (Phase-1) is treated as initially *given*. The model phase is elaborated by the joint expertise and knowledge domain of engineers and technical managers who collaborate in order to define the final modelling of the problem. The modelling is embedded in the objective functions that aim to emulate the physics of the problem as initially apprehended by the engineers and technical managers. When the objective functions are defined, the problem-structure can be revealed by the non-dominated solutions that can be obtained from an MOEA. In this sense, the problem-structure revealed (also referred to as *learnt*) by the non-dominated solutions is able to support the DM objective preference-rankings. The previous statement defines the basis for

this chapter. It is intended to integrate all the three phases (Figure 5.2) so that the link between the modelling (Phase-1) and decision-making (Phase-3) could be exploited and also how that integration could bring *objectivity*, *repeatability*, *consistency* and *coherence* in the DM's preferences.

## 5.2 Proposed Machine Learning Based Framework for Decision Support to the Decision Maker

---

**Framework 2:** Machine learning based framework to facilitate *objectivity*, *repeatability*, *consistency* and *coherence* in DMs' preferences

---

**Input:**

A non-dominated solution set obtained from an MOEA, corresponding to initial objective set  $\mathcal{F}_0 = \{f_1, \dots, f_M\}$ .

- 1 **begin**
  - 2     Compute a positive semi-definite matrix:  $R$  (Equation 2.11) or  $K$  (Equation 2.15), for L-PCA and NL-MVU-PCA, respectively.
  - 3     Compute the eigenvalues and eigenvectors of  $R$  and  $K$  respectively (Section 4.1.2).
  - 4     Perform the eigenvalue analysis (Section 4.1.3) to identify the set of important objectives  $\mathcal{F}_e \subseteq \mathcal{F}_0$ .
  - 5     Perform the RCM analysis (Section 4.1.4) to identify the identically correlated subsets ( $\mathcal{S}$ ) in  $\mathcal{F}_e$ . If there is no such subset,  $\mathcal{F}_s = \mathcal{F}_e$ .
  - 6     Apply the selection scheme (Section 4.1.5) to identify the most significant objective in each  $\mathcal{S}$ , to arrive at  $\mathcal{F}_s$ , such that  $\mathcal{F}_s \subseteq \mathcal{F}_e \subseteq \mathcal{F}_0$ .
  - 7     Compute the error  $\mathcal{E}_i$  associated with each objective using Equation 5.1.
  - 8     I: For revelation of an *essential* objective set: Determine the essential objective set  $\mathcal{F}_s$  and the corresponding error  $\mathcal{E}_{redn}$  using Equation 5.3.
  - 9     II: For preference-ranking of all the objective functions: Determine the weight  $w_i$  associated with each objective using Equation 5.4.
  - 10    III: For  $\delta$ -MOSS analysis: For a DM specified  $\delta$ , determine the  $\delta$ -minimal set  $\mathcal{F}_{\{\delta\}s}$  using Equation 5.5.
  - 11    IV: For  $k$ -EMOSS analysis: For a DM specified  $k$ , determine the  $k$ -minimal error  $\mathcal{E}_k^n$  and the  $k$ -minimal set  $\mathcal{F}_{\{k\}s}$  using Equation 5.6.
  - 12 **end**
- 

The framework basic steps are summarised in Framework 2 and a schematic is provided in Figure 5.3. The general lines of the framework are as follows:

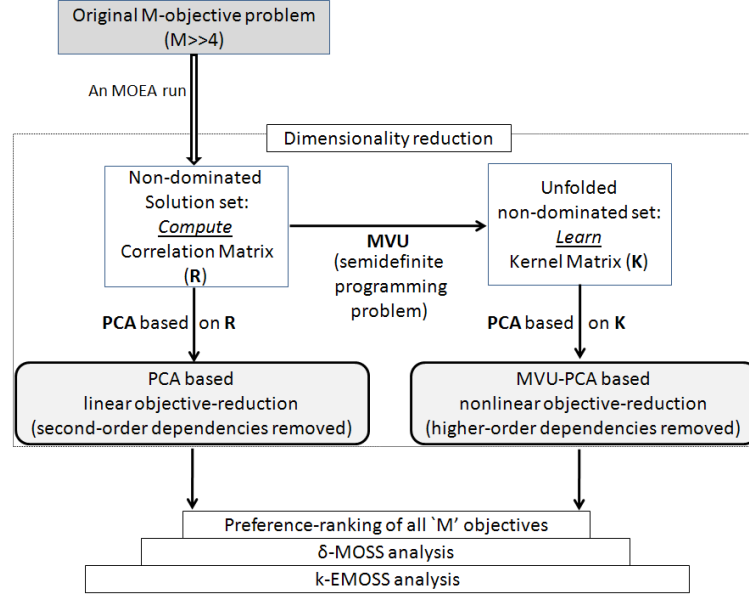


Figure 5.3: A schematic for the proposed machine learning based framework to facilitate *objectivity, repeatability, consistency* and *coherence* in DM's preferences.

1. Operates on the objective vectors of the non-dominated solution sets.
2. Accurate results are expected if the correlation-structure of the given solution set is the same as the one on the POF.
3. Reveals the POF structure by applying machine learning techniques from the following perspectives:

I Revelation of an *essential* objective set: Here, for a given  $M$ -objective problem denoted by  $\mathcal{F}_0 = \{f_1, f_2, \dots, f_M\}$ , the framework reveals an *essential* objective set—the smallest set of conflicting objectives  $(\mathcal{F}_{\mathcal{T}}, |\mathcal{F}_{\mathcal{T}}| = m \ (m \leq M))$  which can generate the same POF as that obtained by the original problem.

II Preference-ranking of all the objective functions: Here, the framework identifies the preference-weight  $(w_i)$  for each objective, such that  $w_i \geq 0$  and  $\sum_{i=1}^M w_i = 1$ . This is achieved by measuring the variance accounted by each objective which can give an indication of the error that can be incurred if an objective were to be eliminated.

- III  $\delta$ -MOSS perspective: The aim is to obtain the smallest objective (sub)set for a specified  $\delta$ -error, where  $0 \leq \delta \leq 1$ . The (sub)set determined is referred to as  $\delta$ -minimal set, denoted by  $\mathcal{F}_{\{\delta\}s}$ , with size  $|\mathcal{F}_{\{\delta\}s}|$ . In this perspective the DM is able to allow some  $\delta$ -error to which he/she is interested to know the smallest (sub)set of possible objectives that incur in an error *just less than or equal to*  $\delta$ . This type of analysis is referred to as  $\delta$ -MOSS ( $\delta$ -Minimum Objective Subset).
- IV  $k$ -EMOSS perspective: The aim is to find an objective (sub)set with a specified size  $k$  that incurs in the minimal possible error. The (sub)set determined is referred to as  $\mathcal{F}_{\{k\}s}$  and the corresponding  $k$ -minimal error is denoted by  $\mathcal{E}_k^n$ . In this perspective the DM is interested to know an objective (sub)set for a specified fraction  $p$  ( $0 \leq p \leq 1$ ) of the original number of objective ( $M$ ) that he/she wishes to retain with size  $k = \lceil pM \rceil$ , which incurs in the minimal possible error. This type of analysis is referred to as  $k$ -EMOSS (Minimum Objective Subset of Size  $k$  with Minimum Error).
- V Visual representation: A simply yet meaningful visual representation of the above analysis is proposed. This could serve as a snap-shot guide for the DMs to base their preferences on.

The steps 2-6 in Framework 2 are identical to steps 3-7 in Framework 1 from Chapter 4. This means that Framework 2 is also capable of finding an essential objective set ( $\mathcal{F}_s$ ) corresponding to the given non-dominated solutions. However, note that the first step in Framework 2 does not involve the generation of a set of non-dominated solutions by an MOEA as Framework 1 did and instead the non-dominated solution set is considered as an input of the framework together with the corresponding initial objective set  $\mathcal{F}_0$ . This is due to the fact that Framework 2 is not meant to run iteratively like Framework 1, where in each iteration a new set of non-dominated solutions is generated and the cardinality of the essential objective set will keep reducing until two consecutive iterations generate the same

set. This means that the focus of Framework 2 is to provide decision support to the DM based only on the provided non-dominated solution set, and therefore, for this purpose there is no need to run this framework in an iterative manner.

The error computation in Framework 1 consists of measuring the variance left unaccounted when objectives in the redundant objective set  $\mathcal{F}_{redn}$  are discarded, as given by Equation 4.4. However, in here the error computation measures the variance left unaccounted when each objective were to be eliminated one at a time. This is a requirement for determining the preference-ranking of all the objective functions, and the  $\delta$ -MOSS and  $k$ -EMOSS analysis. Towards this, consider the following steps:

1. Computation of the error that would be incurred, if each objective were to be eliminated one at a time. The proposed error measure given by Equation 5.1 accounts for the proportion of the variance left unaccounted, if an objective were to be eliminated. The error corresponding to  $f_i \in \mathcal{F}_s$  remains to be the corresponding  $c_i^M$ , while the error for an  $f_i \in \mathcal{F}_{redn} = \mathcal{F}_0 \setminus \mathcal{F}_s$  is scaled down because part of its variance is already accounted for by one of the critical objectives ( $f_i \in \mathcal{F}_s$ ).

$$\left. \begin{aligned} \mathcal{E}_i &= c_i^M (1.0 - \max_{j \in \mathcal{F}_s} \{\delta_{ij} \cdot R_{ij}\}) \text{ for } f_i \in \mathcal{F}_{redn} \\ \mathcal{E}_i &= c_i^M \text{ for } f_i \in \mathcal{F}_s \\ \text{where:} \\ c_i^M &= \sum_{k=1}^M e_k f_{ik}^2 \\ \delta_{ij} &= \begin{cases} 1, & \text{if } f_i \text{ and } f_j \text{ are identically correlated} \\ 0, & \text{otherwise} \end{cases} \\ R_{ij} &= \text{Correlation strength between } f_i \text{ and } f_j \end{aligned} \right\} \quad (5.1)$$

2. Further processing of the  $\mathcal{E}_i$ s in Equation 5.1, as below:

(a) Normalisation of the errors is given by

$$\mathcal{E}_i^n = \mathcal{E}_i / \sum_{j=1}^M \mathcal{E}_j; \quad (5.2)$$

(b) Sorting of the  $\mathcal{E}_i^n$ : Let the  $\mathcal{E}_i^n$ s sorted in *ascending order* be given by  $\mathcal{E}_{r_1}^n$  ( $\mathcal{E}_{r_1}^n \leq \dots \leq \mathcal{E}_{r_M}^n$ ) and the corresponding objective set by  $\{f_{r_1}, \dots, f_{r_M}\}$ . In that,  $f_{r_1} \triangleleft f_{r_2} \triangleleft, \dots, f_{r_{M-1}} \triangleleft f_{r_M}$ , where  $\triangleleft$  denotes *less-important-than* relation.

Given the above pre-processing of the error computation, consider the description of each perspective mentioned in Framework 2 in the following subsections.

### 5.2.1 Revelation of an *Essential* Objective Set

An *essential* objective set is given by  $\mathcal{F}_s$ , and the corresponding error is given by

$$\mathcal{E}_{redn} = \sum_{i|f_i \in \mathcal{F}_{redn}} \mathcal{E}_i, \quad (5.3)$$

which is the error incurred when  $\mathcal{F}_{redn}$  is omitted.

### 5.2.2 Preference-ranking of all the Objectives

The preference-weight for each objective could be given by

$$w_i = \mathcal{E}_i^n = \mathcal{E}_i / \sum_{j=1}^M \mathcal{E}_j \quad (\text{ensuring that } w_i \geq 0 \text{ and } \sum_{i=1}^M w_i = 1), \quad (5.4)$$

and the preference-ranking of all the objectives could be established by the sorted  $\mathcal{E}_i^n$ s (as above). This argument could be justified as follows. Let,  $u$  and  $v$  be two objectives such that  $\mathcal{E}_u^n \gg \mathcal{E}_v^n$ , implying that the error incurred by discarding the objective  $u$  is far greater than the error incurred if objective  $v$  were to be discarded. In other words, for higher accuracy, the objective  $u$  needs to be preferred over  $v$  (or the solutions which are better in  $u$  need to

be preferred over those which are better in  $v$ ).

### 5.2.3 $\delta$ -MOSS Analysis

The  $\delta$ -MOSS analysis can be performed as follows: For a DM specified  $\delta$ , it involves finding the *minimum*  $J$  as given by

$$\begin{aligned} \sum_{j=1}^J \mathcal{E}_{r_j}^n &\geq \delta, \\ \text{such that the } \delta\text{-minimal set is given by:} \\ \mathcal{F}_{\{\delta\}s} &= \{f_{r_J}, f_{r_{J+1}}, \dots, f_{r_M}\}. \end{aligned} \tag{5.5}$$

### 5.2.4 $k$ -EMOSS Analysis

The  $k$ -EMOSS analysis can be conducted as follows: For a DM specified  $k$ , it involves determining the  $k$ -minimal error given by

$$\begin{aligned} \mathcal{E}_k^n &= \sum_{j=1}^{M-k} \mathcal{E}_{r_j}^n, \\ \text{corresponding to which the } k\text{-minimal set is given by:} \\ \mathcal{F}_{\{k\}s} &= \{f_{r_{M-k+1}}, f_{r_{M-k+2}}, \dots, f_{r_M}\}. \end{aligned} \tag{5.6}$$

Given that the proposed framework provides a decision support through four perspectives as above (to be backed by a visual representation of the analysis results), and given that this framework could be applied multiple times during an MOEA run, in here it is envisaged that the framework can facilitate with *objectivity*, *repeatability*, *consistency*, and *coherence* in the DM's preferences.

## 5.3 Experimental Settings

In Chapter 4, the main focus has been on demonstrating the accuracy of the proposed machine learning based objective reduction algorithms in determining the 0-minimum objective



set for solutions sets characterised by: (i) *unnoised* signal (generated on the true POF and denoted by  $\mathcal{N}_{\mathcal{P}}$ ), and (ii) with different degrees of *noise* (generated by NSGA-II and denoted by  $\mathcal{N}_{\mathcal{NS}}$  and generated by  $\epsilon$ -MOEA and denoted by  $\mathcal{N}_{\epsilon}$ ). This has been shown on: (a) redundant test problems, such as DTLZ5( $I, M$ ) and WFG3( $M$ ); (b) non-redundant test problem, such as DTLZ1( $M$ ), DTLZ2( $M$ ), DTLZ3( $M$ ) and DTLZ4( $M$ ); and (c) two real-world problems, namely, multi-speed gearbox design and storm drainage system. Besides this, a comparative analysis between the proposed algorithms and an alternative approach based on dominance relation preservation has also been conducted.

In this chapter the main focus is on demonstrating how the proposed framework can be utilised as a decision support for the DM in articulating his/her preferences with rationality. Therefore, most of the experimental results focuses on real-world problems since the usefulness and practicality of the decision support can be better demonstrated when the objective functions have an actual physical meaning. This leads to the application of the framework to four real-world problems that have been identified in the literature and whose details will be provided along Sections 5.4–5.7. Despite this, it is also important to investigate the efficacy of the proposed framework in terms of  $\delta$ -MOSS and  $k$ -EMOSS analysis when the test problems characteristics are known. This analysis is more useful when applied to redundant test problems and with different degrees of nonlinearity, where the *nonlinearity* relates to the nonlinear distribution (in objective space) of the non-dominated solutions (further details will be provided in Section 5.8). Therefore, different versions of the test problems DTLZ5( $I, M$ ) and DTLZ2<sub>BZ</sub> [2], are considered. DTLZ2<sub>BZ</sub> is chosen since different objectives have different degrees of nonlinearity, and as discussed in [2], this problem resembles the real-world problems from two perspectives: (i) the different objectives have different scales, and (ii) with the elimination of a subset of objectives, the POF does not collapse to a single solutions.

Moreover, non-dominated solutions generated for the above problems and the corresponding MOEA parameters are as follows:

- For real-world problems with the exception of the radar waveform problem (Section 5.4): 20 runs are conducted using NSGA-II corresponding to 20 equally spaced seed. Each run consists of 2000 generations and the final 200 non-dominated solutions are denoted by  $\mathcal{N}_{\mathcal{NS}}$ . Other NSGA-II parameters are: (i) 0.9 and 0.1 for probability of crossover and mutation, respectively; and (ii) 5 and 20 for distribution index for crossover and mutation, respectively.
- for test problems:
  - solutions generated on the true POF (denoted by  $\mathcal{N}_{\mathcal{P}}$ ). Generated using the sampling scheme in [44]. The number of solutions is given by  $N = C_{m-1}^{H+m-1}$ , where  $H$  is a user defined parameter and  $m$  corresponds to the dimensions of the POF. In here, for  $m \leq 10$ ,  $H$  is chosen such that the number of solutions is greater than or equal to 200, and  $H = 4$  for  $m > 10$ .
  - solutions generated by  $\epsilon$ -MOEA (denoted by  $\mathcal{N}_{\epsilon}$ ). The settings are: (i) population size is 200, (ii) number of generations is 2000, (iii) probability of crossover is 0.9, (iv) probability of mutation is 0.1, (v) distribution index for crossover is 5, (vi) distribution index for mutation is 20, and (vii)  $\epsilon = 0.3$ .

Notably, the above parameters selected to generate the non-dominated solutions have also been adopted in Chapter 4.

## 5.4 Experimental Results on Radar Waveform Problem

This section demonstrates the application of the proposed decision support framework on a radar waveform problem, and compares the results with those reported by [2]. To this end, the NL-MVU-PCA (referred to as Approach-A) and the exact algorithm from DRP (referred to as Approach-B) have been chosen. For Approach-A, NL-MVU-PCA is chosen

over L-PCA due to its superiority in terms of handling datasets with different distributions (Section 4.14.1). For Approach-B, the exact is chosen over the greedy algorithm since the latter does not guarantee a 0-minimum objective set.

The radar waveform problem, originally described by [162], deals with the design of a waveform for a pulsed Doppler radar, typical of many airborne fighter radar systems, where the aim is to unambiguously measure the range and velocity of the targets that may travel at very high velocities (Mach 5 possible<sup>1</sup>) and may be located at very long distances (100 nautical miles typical). While different versions of this problem exist, depending on the number of variables (pulse repetition interval values), here the analysis is based on the nine-objective, eight-variable version considered by [162], and the non-dominated solution sets taken from [163], correspond to 30 runs of MSOPS-II ([43]), each with 20000 function evaluations and size ranging from 8839 to 9716 points.

The physical meaning of each objective is defined in Equation 5.7, where: (i)  $f_1$  to  $f_8$  are to be maximised, while  $f_9$  is to be minimised, and (ii) the objectives associated with the performance in range, namely 1&3, 5&7, tend to have a degree of correlation, just as the objectives associated with the performance in velocity, namely 2&4, 6&8, do. However, the objectives related with the range and velocity do conflict ([162]).

$$\left. \begin{array}{ll} f_1/f_2 & \equiv \text{Median range/velocity extent of target before schedule is not decodable,} \\ f_3/f_4 & \equiv \text{Median range/velocity extent of target before schedule has blind regions,} \\ f_5/f_6 & \equiv \text{Minimum range/velocity extent of target before schedule is not decodable,} \\ f_7/f_8 & \equiv \text{Minimum range/velocity extent of target before schedule has blind regions,} \\ f_9 & \equiv \text{Time required to transmit total waveform.} \end{array} \right\} \quad (5.7)$$

---

<sup>1</sup>Mach 5 is considered an hypersonic speed that corresponds to 1701.45 m/s, one example is the Boeing X-51 WaveRider which is cooperative effort of the United States Air Force, DARPA, NASA, Boeing, and Pratt & Whitney Rocketdyne.

### 5.4.1 Demonstration of the Approach-A (NL-MVU-PCA based Framework)

This section demonstrates the application of Approach-A, corresponding to one of the 30 solution sets, detailed above. For brevity, these solution sets originally obtained from MSOPS-II are referred to as  $\mathcal{N}_{MS}$ .

#### Perspective-I: Revelation of an *Essential* Objective Set

The application of Approach-A is detailed in Table 5.1 and the steps are:

1. The correlation matrix ( $R$ ) and the eigenvalues/eigenvectors of matrix  $K$  are shown in Tables 5.1a and 5.1b, respectively.
2. The eigenvalue analysis shown in Table 5.1b reveals that: (i) the number of *significant* principal components ( $N^v$ ) is four (necessary to meet  $\theta = 0.997$ ), and (ii)  $\mathcal{F}_e = \mathcal{F}_0$ , as no objective could be eliminated on account of non-conflict along the *significant* principal components.
3. The RCM analysis shown in Table 5.1c reveals: (i) nine *potentially identically* correlated objective subsets, with only two distinct sets of composition— $\{f_1, f_3, f_5, f_7\}$  comprising all the range related objectives, and  $\{f_2, f_4, f_6, f_8, f_9\}$  comprising all the velocity related objectives along with the objective relating to transmission time, (ii) a relatively low  $T_{cor} = 0.371$  indicative of redundancy in the problem, and (iii) nine *identically* correlated objective subsets, with four distinct sets of composition. Unlike  $\hat{\mathcal{S}}_3$ ,  $\mathcal{S}_3$  does not contain  $f_7$ , since  $R_{37} = 0.286 < T_{cor}$ . The same explains the difference in the compositions of  $\hat{\mathcal{S}}_7$  and  $\mathcal{S}_7$ .
4. Based on the selection score ( $sc_i$ ) already computed in Table 5.1b, the following objectives are picked as *essential*: (i)  $f_3$  from subsets  $\mathcal{S}_1$ ,  $\mathcal{S}_3$ , and  $\mathcal{S}_5$ ; (ii)  $f_4$  from subsets

Table 5.1: Radar waveform problem: Approach-A (NL-MVU-PCA based framework), corresponding to  $\mathcal{N}_{MS}$  (one run)

(a) Correlation matrix ( $R$ )

	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$f_7$	$f_8$	$f_9$
$f_1$	1.000	-0.872	0.863	-0.916	0.593	-0.412	0.522	-0.744	-0.925
$f_2$	-0.872	1.000	-0.939	0.954	-0.516	0.485	-0.365	0.865	0.953
$f_3$	0.863	-0.939	1.000	-0.942	0.514	-0.471	0.286	-0.855	-0.984
$f_4$	-0.916	0.954	-0.942	1.000	-0.560	0.469	-0.441	0.823	0.970
$f_5$	0.593	-0.516	0.514	-0.560	1.000	-0.197	0.487	-0.406	-0.567
$f_6$	-0.412	0.485	-0.471	0.469	-0.197	1.000	-0.121	0.558	0.465
$f_7$	0.522	-0.365	0.286	-0.441	0.487	-0.121	1.000	-0.150	-0.388
$f_8$	-0.744	0.865	-0.855	0.823	-0.406	0.558	-0.150	1.000	0.843
$f_9$	-0.925	0.953	-0.984	0.970	-0.567	0.465	-0.388	0.843	1.000

(b) Eigen-decomposition, eigenvalue analysis and selection scores

	$e_1 = 0.809$	$e_2 = 0.163$	$e_3 = 0.019$	$e_4 = 0.009$	
	$V_1$	$V_2$	$V_3$	$V_4$	$sc_i$
$f_1$	-0.170	0.109	<u>-0.916</u>	<u>-0.093</u>	0.173
$f_2$	<u>0.181</u>	<u>-0.162</u>	<u>0.081</u>	<u>-0.147</u>	0.176
$f_3$	<u>-0.888</u>	<u>-0.157</u>	<u>0.270</u>	<u>-0.026</u>	0.750
$f_4$	<u>0.187</u>	<u>-0.165</u>	<u>0.086</u>	<u>-0.148</u>	0.181
$f_5$	<u>0.105</u>	<u>-0.048</u>	<u>0.000</u>	<u>0.935</u>	0.101
$f_6$	<u>0.179</u>	<u>-0.161</u>	<u>0.080</u>	<u>-0.147</u>	0.174
$f_7$	<u>0.036</u>	<u>0.910</u>	<u>0.231</u>	<u>-0.075</u>	0.182
$f_8$	<u>0.183</u>	<u>-0.160</u>	<u>0.082</u>	<u>-0.147</u>	0.177
$f_9$	<u>0.185</u>	<u>-0.163</u>	<u>0.083</u>	<u>-0.149</u>	0.179

(c) RCM analysis

Potential identically correlated set	Eq 4.1(i)	$\hat{\mathcal{S}}_1 = \hat{\mathcal{S}}_3 = \hat{\mathcal{S}}_5 = \hat{\mathcal{S}}_7 = \{f_1, f_3, f_5, f_7\}$ $\hat{\mathcal{S}}_2 = \hat{\mathcal{S}}_4 = \hat{\mathcal{S}}_6 = \hat{\mathcal{S}}_8 = \hat{\mathcal{S}}_9 = \{f_2, f_4, f_6, f_8, f_9\}$
Correlation threshold: $T_{cor}$	Eq 4.1(iii)	$1.0 - 0.808941 \times (1.0 - 2/9) = 0.370824$
Identically correlated set	Eq 4.1(ii)	$\mathcal{S}_1 = \mathcal{S}_5 = \{f_1, f_3, f_5, f_7\}$ $\mathcal{S}_3 = \{f_1, f_3, f_5\}; \mathcal{S}_7 = \{f_1, f_5, f_7\}$ $\mathcal{S}_2 = \mathcal{S}_4 = \mathcal{S}_6 = \mathcal{S}_8 = \mathcal{S}_9 = \{f_2, f_4, f_6, f_8, f_9\}$

(d) Error computation

(e) Error based objective-ranking

Category	With reference to Equation 5.1				Ascending order		Cumulative
	$c_i^M$	$\max\{\delta_{i,j}, R_{ij}\}$	$\mathcal{E}_i$	$\mathcal{E}_i^n$			$\mathcal{E}_i^n$
$f_1 \in \mathcal{F}_{redn}$	0.041116	0.863848 ( $R_{13}$ )	0.005597	0.006558	$f_{r_1}$	$f_9$	0.001107
$f_2 \in \mathcal{F}_{redn}$	0.031403	0.954665 ( $R_{24}$ )	0.001423	0.001667	$f_{r_2}$	$f_2$	0.002775
$f_3 \in \mathcal{F}_s$	0.644745	—	0.644744	0.755382	$f_{r_3}$	$f_1$	0.009334
$f_4 \in \mathcal{F}_s$	0.033398	—	0.033397	0.039128	$f_{r_4}$	$f_8$	0.015940
$f_5 \in \mathcal{F}_{redn}$	0.017168	0.514645 ( $R_{53}$ )	0.008332	0.009762	$f_{r_5}$	$f_5$	0.025703
$f_6 \in \mathcal{F}_{redn}$	0.030609	0.469935 ( $R_{64}$ )	0.016224	0.019008	$f_{r_6}$	$f_6$	0.044712
$f_7 \in \mathcal{F}_s$	0.137228	—	0.137228	0.160776	$f_{r_7}$	$f_4$	0.083840
$f_8 \in \mathcal{F}_{redn}$	0.031862	0.823009 ( $R_{84}$ )	0.005639	0.006606	$f_{r_8}$	$f_7$	0.244617
$f_9 \in \mathcal{F}_{redn}$	0.032472	0.970888 ( $R_{94}$ )	0.000945	0.001107	$f_{r_9}$	$f_3$	1.000000

$\mathcal{S}_2, \mathcal{S}_4, \mathcal{S}_6, \mathcal{S}_8$  and  $\mathcal{S}_9$ ; (iii)  $f_7$  from subset  $\mathcal{S}_7$ . As a result,  $\mathcal{F}_s = \{f_3, f_4, f_7\}$ , while  $\mathcal{F}_{redn} = \{f_1, f_2, f_5, f_6, f_8, f_9\}$ .

5. The error corresponding to elimination of  $\mathcal{F}_{redn}$  (based on Equation 5.3 and Table 5.1d) is given by

$$\mathcal{E}_{redn} = \sum_{i \in \{1,2,5,6,8,9\}} \mathcal{E}_i = 3.816349\%. \quad (5.8)$$

### Perspective-II: Preference-ranking of all the Objectives

Based on Equation 5.4, the preference-weights for each objective ( $\mathcal{E}_i^n$ ), are shown in Table 5.1d. Sorting of these preference-weights establishes the preference-ranking of all the objectives (Table 5.1e), as follows

$$f_3 \triangleright f_7 \triangleright f_4 \triangleright f_6 \triangleright f_5 \triangleright f_8 \triangleright f_1 \triangleright f_2 \triangleright f_9, \quad (5.9)$$

where  $\triangleright$  denotes *more-important-than* relation.

### Perspective-III: $\delta$ -MOSS Analysis

For the  $\delta$ -MOSS analysis, the objective normalised errors ( $\mathcal{E}_i$ ) obtained in Table 5.1d are sorted in ascending order (denoted as  $\mathcal{E}_i^n$ ) and their cumulative is computed in Table 5.1e. With the information provided in Table 5.1e together with Equation 5.5, the DM is able to know what is the set of minimal number of objectives that incur in a error less than or equal to a specified  $\delta$ . If  $\delta = 0.1$  is specified a  $\delta$ -minimal objective set (denoted  $\mathcal{F}_{\{0.1\}^s}$ ) can be obtained so that the number of discarded objectives will not exceed an error of 10%. For the problem in consideration and the information provided in Table 5.1e the  $\delta$ -minimal objective sets for different  $\delta$ 's are as follows:

- $\mathcal{F}_{\{\delta=0.1=0.2\}^s} = \{f_3, f_7\},$

- $\mathcal{F}_{\{\delta > 0.2446\}s} = \{f_3\}$ .

Note that the  $\delta$ -minimal objective set for  $\delta = 0.1$  and  $\delta = 0.2$  is  $\{f_3, f_7\}$  for both cases, while for  $\delta > 0.2446$  the same  $\delta$ -minimal objective set is obtained, that is,  $\{f_3\}$ .

#### Perspective-IV: $k$ -EMOSS Analysis

For the  $k$ -EMOSS analysis the DM is able to specify a  $p\%$  of the original number of objectives, i.e.,  $k$  objectives, where  $k = \lceil pM \rceil$  that he/she may want to retain. This analysis provides a  $k$ -minimal objective set with corresponding  $k$ -minimal error so that further objective set reductions inform the DM of the error that they will incur. In this problem, and for  $\delta$ -MOSS, the analysis is conducted using Table 5.1e but with Equation 5.6. For different  $k$ 's some  $k$ -minimal objective sets with corresponding  $k$ -minimal errors are the following:

- $p = 10\%$ , i.e.,  $k = \lceil 0.10 \times 9 \rceil = 1$ :  $\mathcal{F}_{\{k=1\}s} = \{f_3\}$  and  $\mathcal{E}_{k=1}^n = 0.244617 \equiv 24.4617\%$ ,
- $p = 20\%$ , i.e.,  $k = \lceil 0.20 \times 9 \rceil = 2$ :  $\mathcal{F}_{\{k=2\}s} = \{f_3, f_7\}$  and  $\mathcal{E}_{k=2}^n = 0.083840 \equiv 8.3840\%$ ,
- $p = 30\%$ , i.e.,  $k = \lceil 0.30 \times 9 \rceil = 3$ :  $\mathcal{F}_{\{k=3\}s} = \{f_3, f_4, f_7\}$  and  $\mathcal{E}_{k=3}^n = 0.044712 \equiv 4.4712\%$ .

#### 5.4.2 Sensitivity of the Decision Support to Different Samples of

$$\mathcal{N}_{\mathcal{MS}}$$

This section investigates the sensitivity of the decision support offered by the proposed framework, to different samples of  $\mathcal{N}_{\mathcal{MS}}$ . The framework is repeated for the 30 samples of  $\mathcal{N}_{\mathcal{MS}}$ , available at [162], and the mean and standard deviation for  $|\mathcal{F}_{\{\delta\}}|$  and  $\mathcal{E}_k^n$  are reported in Tables 5.2 and 5.3, for  $\delta$ -MOSS and  $k$ -EMOSS analysis, respectively. It can be seen that the mean values for  $|\mathcal{F}_{\{\delta\}}|$  and  $\mathcal{E}_k^n$  (in respective cases), strongly comply with those reported by the  $\delta$ -MOSS and  $k$ -EMOSS analysis provided above, and the standard deviation is not too significant. Hence, it can be inferred that the proposed framework is fairly robust and thus the resulting decision support can be relied upon.

Table 5.2: Radar waveform problem:  $\delta$ -MOSS analysis by the Approach-A (NL-MVU-PCA based framework), averaged over 30  $\mathcal{N}_{\mathcal{MS}}$ 

DM Input	Frequency of occurrence of objectives									Output Indicator
$\delta$ (%)	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$f_7$	$f_8$	$f_9$	$ \mathcal{F}_{\{\delta\}} $
00	08	00	30	30	00	00	14	01	00	$2.77 \pm 0.50$
10	08	00	30	00	00	00	30	00	00	$2.27 \pm 0.45$
20	00	00	30	00	00	00	19	00	00	$1.63 \pm 0.49$
30	00	00	30	00	00	00	00	00	00	$1.00 \pm 0.00$
40	00	00	30	00	00	00	00	00	00	$1.00 \pm 0.00$
50	00	00	30	00	00	00	00	00	00	$1.00 \pm 0.00$
60	00	00	30	00	00	00	00	00	00	$1.00 \pm 0.00$
70	00	00	30	00	00	00	00	00	00	$1.00 \pm 0.00$
80	00	00	30	00	00	00	00	00	00	$1.00 \pm 0.00$
90	00	00	30	00	00	00	00	00	00	$1.00 \pm 0.00$

Table 5.3: Radar waveform problem:  $k$ -EMOSS analysis by the Approach-A (NL-MVU-PCA based framework), averaged over 30  $\mathcal{N}_{\mathcal{MS}}$ 

DM Input	Frequency of occurrence of objectives									Output Indicator
$k$	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$f_7$	$f_8$	$f_9$	$\mathcal{E}_k^n$
1	00	00	30	00	00	00	00	00	00	$0.217 \pm 0.025$
2	00	00	30	00	00	00	30	00	00	$0.099 \pm 0.020$
3	08	00	30	22	00	00	30	00	00	$0.056 \pm 0.015$
4	08	00	30	30	00	21	30	01	00	$0.030 \pm 0.005$
5	11	00	30	30	17	30	30	02	00	$0.018 \pm 0.001$
6	19	00	30	30	27	30	30	14	00	$0.010 \pm 0.001$
7	30	00	30	30	30	30	30	30	00	$0.003 \pm 0.000$
8	30	30	30	30	30	30	30	30	00	$0.001 \pm 0.000$



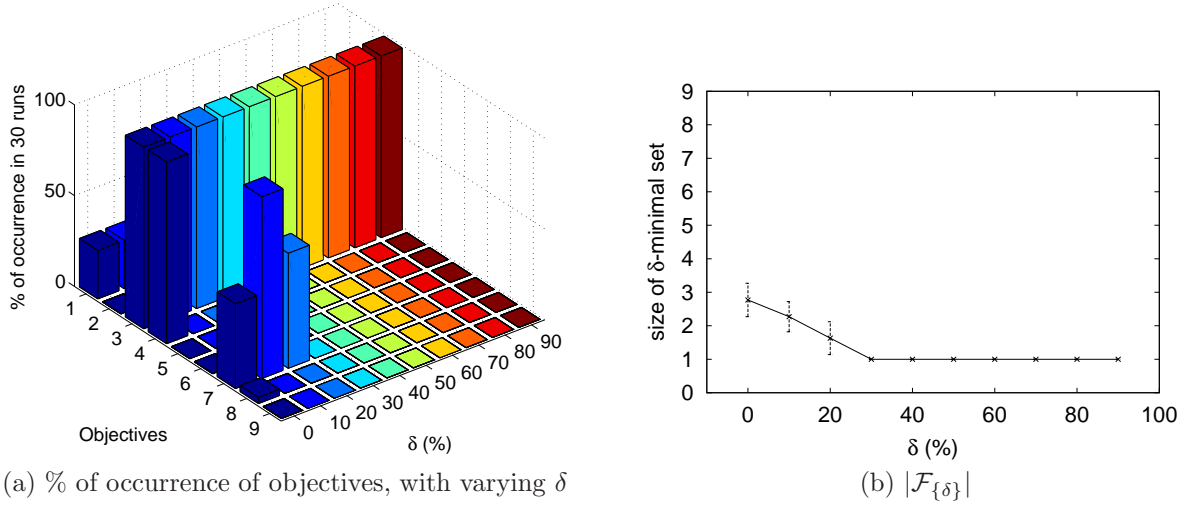


Figure 5.4: Radar waveform problem: Visual representation of the  $\delta$ -MOSS analysis by the Approach-A (NL-MVU-PCA based framework), averaged over 30  $\mathcal{N}_{MS}$

### 5.4.3 Visual Representation of the Results

It is important that the decision support be offered in a manner that the information exchange with the DMs is uncomplicated and time-efficient. Thus a simple, yet effective visualisation scheme is proposed, as presented in: (i) Figures 5.4a and 5.4b in the case of  $\delta$ -MOSS analysis, and (ii) Figures 5.5a and 5.4b in the case of  $k$ -EMOSS analysis.

### 5.4.4 Key Highlights of the Decision Support Offered by the Proposed Framework

The key highlights of the proposed decision support offered by the Approach-A, are as follows:

- At the expense of an error of 3.816349%, the original nine-objective problem could be reduced to a three-objective problem, where the *essential* set is given by  $\mathcal{F}_s = \{f_3, f_4, f_7\}$ . The objectives identified as essential relate to median range ( $f_3$ ), median velocity ( $f_4$ ), and minimum range ( $f_7$ ) extent of target before schedule has blind regions.
- The preference-ranking of all the objectives, given by  $f_3 \triangleright f_7 \triangleright f_4 \triangleright f_6 \triangleright f_5 \triangleright f_8 \triangleright f_1 \triangleright f_2 \triangleright f_9$ ,

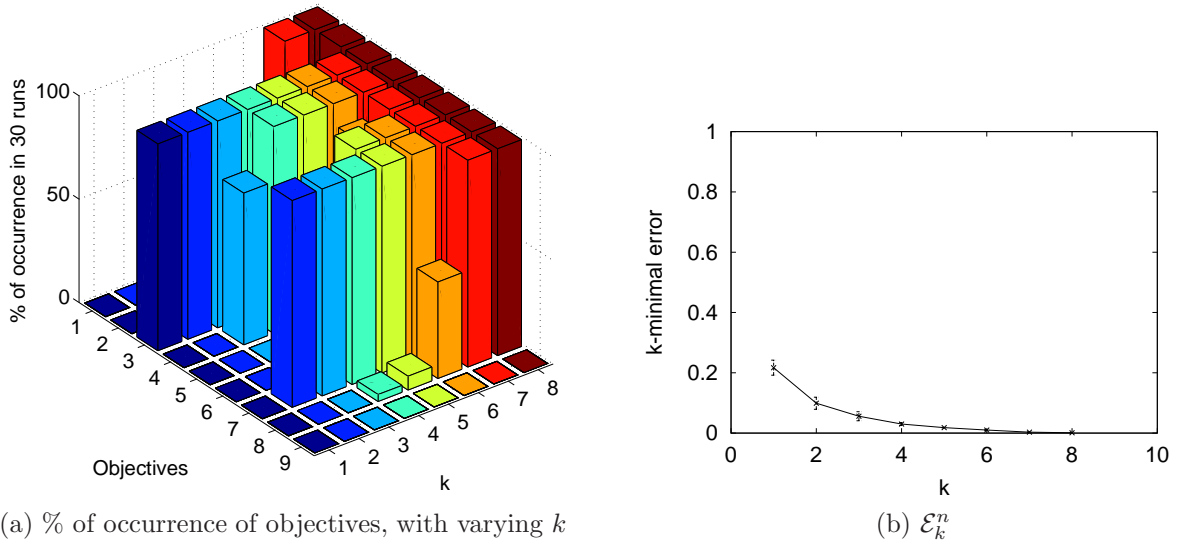


Figure 5.5: Radar waveform problem: Visual representation of the  $k$ -EMOSS analysis by the Approach-A (NL-MVU-PCA based framework), averaged over 30  $\mathcal{N}_{MS}$

where  $\triangleright$  denotes *more-important-than* relation, could enable the DM to establish a total or partial order on the presented solutions, which otherwise is a difficult task, particularly when many objectives are involved.

- If the DM allows for an error greater than 24.4617%, then accounting for only a single objective, namely,  $f_3$  will suffice. On the other hand, if the DM wants to account for only two objectives which correspond to minimum error, then  $f_3$  and  $f_7$  need to be accounted for.
- The proposed visualisation scheme could expose the preference-ranking of all the objectives to the DM, in one snap-shot.

#### 5.4.5 Comparative Analysis with the Approach-B (DRP Based Exact Algorithm)

Based on the above results and those revealed by [2] for Approach-B, the following comparisons can be made:

1. Computational complexity and its implications on the solution set size that could be

handled: As noted in Section 4.14.2, the computational complexity of the objective reduction approaches is:

- (a) For DRP-based algorithms, namely exact (denoted by Approach-B) and greedy, is  $O(N^2 M 2^M)$  and  $O(\min\{N^2 M^3, N^4 M^2\})$ , respectively.
- (b) For PCA-based algorithms, namely NL-MVU-PCA (denoted by Approach-A) and L-PCA, is  $O(M^6)$  (worst case) and  $O(NM^2 + M^3)$ , respectively.

The fact that the computational complexity of the DRP-based exact algorithm is exponential in  $M$  and quadratic in  $N$ , puts stringent limitations on the number of objectives and the solution set size that could be handled in a reasonable computational time. Though, the limitations on the DRP-based greedy algorithm are less stringent, they are still likely to be stronger than in the case of PCA-based algorithms (because in MaOPs, generally  $N \gg M$ ). The implications of the above observation in the context of the radar waveform problem are the following:

- The proposed framework based on Approach-A could efficiently handle the 30 samples of  $\mathcal{N}_{\mathcal{MS}}$ , available at [163], where the size of each sample ranged from 8839 to 9716 solutions.
  - The solution set size in the case of Approach-B was reduced to 107, towards which  $\epsilon$ -non-dominated solutions had to be computed, for an arbitrarily chosen error  $\epsilon = 6.2\%$  (page 153 in [2]).
2. Compliance with the trends predicted by [162]: The aforementioned source for the problem, predicts the tendency of  $f_1 \& f_3$ ,  $f_5 \& f_7$  to be correlated, and the same for  $f_2 \& f_4$ ,  $f_6 \& f_8$ . While the Approach-A complies with this prediction, Approach-B show contrasting results. This is affirmed by the following observations:
- In the case of the proposed framework, Tables 5.1a and 5.1c reveal that:

- All the four range related objectives, namely,  $f_1$ ,  $f_3$ ,  $f_5$ , and  $f_7$  are correlated, and the correlation between objectives relating to ‘median’ range ( $R_{13} = 0.863$ ) is stronger than the correlation between objective pairs where one objective relates to ‘median’ and the other to ‘minimum’ range ( $R_{15} = 0.593$ ,  $R_{17} = 0.522$ ,  $R_{35} = 0.514$ , and  $R_{37} = 0.286$ ).
- All the four velocity related objectives, namely,  $f_2$ ,  $f_4$ ,  $f_6$ , and  $f_8$ , are correlated, and the correlation between objectives relating to ‘median’ velocity ( $R_{24} = 0.954$ ) is stronger than the correlation between objective pairs where one objective relates to ‘median’ and the other to ‘minimum’ velocity ( $R_{26} = 0.485$ ,  $R_{28} = 0.865$ ,  $R_{46} = 0.469$ , and  $R_{48} = 0.823$ ).
- The objective relating to minimisation of transmission time ( $f_9$ ) is correlated with the objectives of maximisation of velocity ( $f_2$ ,  $f_4$ ,  $f_6$ , and  $f_8$ ).
- In contrast to the predicted trends, Approach-B reports<sup>2</sup> very high errors between different objective pairs comprising of objectives that are supposed to be correlated. In that, the reported  $\delta$  error is as follows: (i)  $\delta(f_1, f_3) : 0.70 < \delta < 0.75$ , (ii)  $\delta(f_5, f_7) : \delta > 0.80$ , (iii)  $\delta(f_2, f_4) : \delta > 0.80$ , and (iv)  $\delta(f_6, f_8) : 0.70 < \delta < 0.75$ .

3. Reliability of the results: In terms of the terminology adopted in this chapter:

- Approach-A reports  $\mathcal{E}_{redn} = 3.8\%$ , for  $\mathcal{F}_{redn} = \{f_1, f_2, f_5, f_6, f_8, f_9\}$ .
- Approach-B reports  $\mathcal{E}_{redn} = 6.2\%$ , corresponding to the removal of two objectives (that have not been specified).

However, as Approach-B has revealed contrasting trends to those predicted, it is fair to infer the results obtained by the proposed framework as more reliable.

---

<sup>2</sup>This can be observed from Figure 7, Page 154, in [2].

## 5.5 Experimental Results on Car Side-Impact Problem

This section demonstrates the application of the proposed framework on a car side-impact problem. Likewise in Section 5.4, NL-MVU-PCA is referred to as Approach-A.

The car side-impact problem consist in applying a side impact to a car based on European enhanced vehicle-safety committee (EEVC) procedures. The effect of the side and impact on both the dummy and the car is taken into account. The main objective of this problem is to reduce the weight of the car in such as way that 10 constraints are not violated. It is intuitive that reducing the weight of a car improves the fuel economy and vehicle production costs but it may harm safety. In this sense, there is a need to produce a design that better balances weight and safety performance. The problem formulation can be found in [164] and in Equation 5.10 the physical meaning of the objective and constraints is described.

$$\begin{array}{ll}
 \text{Minimise} & f(\mathbf{x}) \equiv \text{Weight}, \\
 (x_1, \dots, x_7) & \\
 \text{Subject to:} & \\
 g_1(\mathbf{x}) \equiv \text{Abdomen load} \leq 1 \text{ kN}, & \\
 g_2(\mathbf{x}) \equiv \text{Viscous criterion: Upper} \leq 0.32 \text{ m/s}, & \\
 g_3(\mathbf{x}) \equiv \text{Viscous criterion: Middle} \leq 0.32 \text{ m/s}, & \\
 g_4(\mathbf{x}) \equiv \text{Viscous criterion: Lower} \leq 0.32 \text{ m/s}, & \\
 g_5(\mathbf{x}) \equiv \text{Upper rib deflection} \leq 32 \text{ mm}, & \\
 g_6(\mathbf{x}) \equiv \text{Middle rib deflection} \leq 32 \text{ mm}, & \\
 g_7(\mathbf{x}) \equiv \text{Lower rib deflection} \leq 32 \text{ mm}, & \\
 g_8(\mathbf{x}) \equiv \text{Pubic force} \leq 4 \text{ kN}, & \\
 g_9(\mathbf{x}) \equiv \text{Vel. of V-Pillar at middle point} \leq 9.9 \text{ mm/ms}, & \\
 g_{10}(\mathbf{x}) \equiv \text{Vel. of front door at V-Pillar} \leq 15.7 \text{ mm/ms}. & 
 \end{array} \quad (5.10)$$

It may be noted that each constraint ( $g_i$ ) represents an upper bound for either the damage to the dummy or the car. Hence, treating each  $g_i$  as an objective (to be minimised), will potentially help improve the safety performance. The resulting eleven-objective problem is

given by Equation 5.11.

$$\left. \begin{array}{ll} \text{Minimise} & F_i(\mathbf{x}) \quad \forall i = 1 \dots 11, \\ (x_1, \dots, x_7) & \\ & g_j(\mathbf{x}) \leq 0 \quad \forall j = 1 \dots 10, \\ \text{where:} & \\ & \left. \begin{array}{l} F_1(\mathbf{x}) = f(\mathbf{x}), \\ F_i(\mathbf{x}) = g_{i-1}(\mathbf{x}) \quad \forall i = 2 \dots 11. \end{array} \right\} \text{Refer Equation 5.10} \end{array} \right\} \quad (5.11)$$

### 5.5.1 Demonstration of the Approach-A (NL-MVU-PCA based Framework)

The analysis that follows corresponds to the application of the Approach-A to one of the 20 non-dominated solution sets.

#### Perspective-I: Revelation of an *Essential* Objective Set

The application of Approach-A is detailed in Tables 5.4 and 5.5. The steps are:

1. The correlation of matrix  $R$  and the eigenvalues and eigenvectors of matrix  $K$  are shown in Table 5.4a and 5.4b, respectively.
2. The eigenvalues analysis is shown in Table 5.4b and the number of principal components ( $N_v$ ) needed to meet the variance threshold ( $\theta = 0.997$ ) is four. In those four principal components all the 11 objectives are selected and therefore no objective reduction is performed at this stage.
3. The RCM analysis in Table 5.4c reveals eight *potentially identically* correlated objective subsets, with three distinct sets of composition— $\{f_2, f_5, f_9\}$ ,  $\{f_3, f_4, f_6\}$ , and  $\{f_8, f_{10}\}$ . Due to a relatively low  $T_{cor} = 0.518$  indicative of redundancy in the problem, the eight objective subsets are considered *identically* correlated since the correlations are:

- In  $\hat{\mathcal{S}}_2$ ,  $\hat{\mathcal{S}}_5$  and  $\hat{\mathcal{S}}_9$ :  $R_{25} = 0.787$ ,  $R_{29} = 0.857$  and  $R_{59} = 0.793$ .

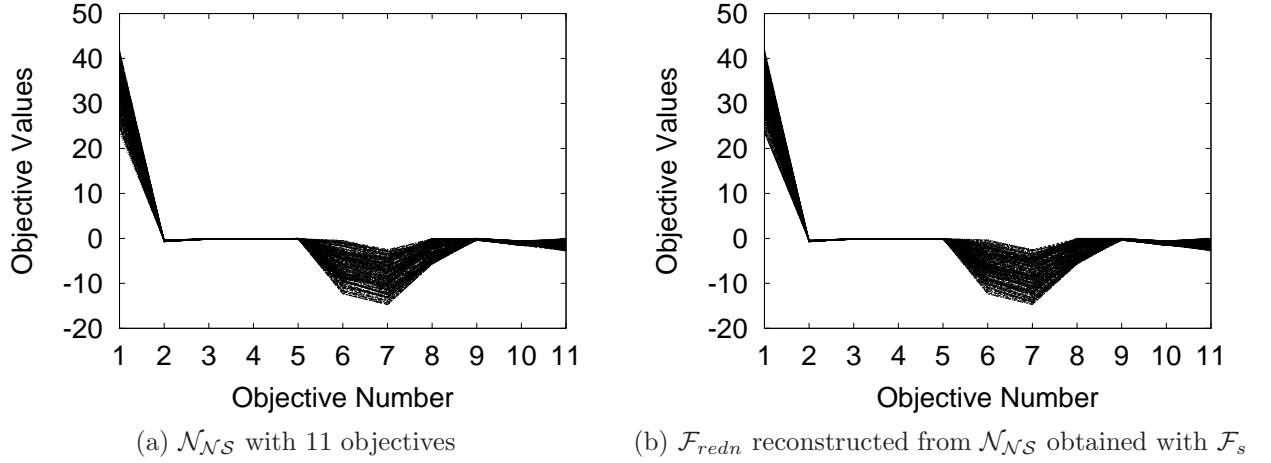


Figure 5.6: Car side-impact problem: Parallel coordinate plots, corresponding to one run of NSGA-II.

- In  $\hat{\mathcal{S}}_3$ ,  $\hat{\mathcal{S}}_4$  and  $\hat{\mathcal{S}}_6$ :  $R_{34} = 0.976$ ,  $R_{36} = 0.941$  and  $R_{46} = 0.977$ .
  - In  $\hat{\mathcal{S}}_8$  and  $\hat{\mathcal{S}}_{10}$ :  $R_{8,10} = 0.990$ .
4. The selection score ( $sc_i$ ) computed for each objective in Table 5.4b allows the scheme to select the most significant objectives from each identically correlated subset  $\mathcal{S}$ . In that,  $f_9$  is selected from  $\mathcal{S}_2$ ,  $\mathcal{S}_5$  and  $\mathcal{S}_9$ ;  $f_6$  from  $\mathcal{S}_3$ ,  $\mathcal{S}_4$  and  $\mathcal{S}_6$ ; and  $f_8$  from  $\mathcal{S}_8$  and  $\mathcal{S}_{10}$ . After this selection the essential objective set is given by  $\mathcal{F}_s = \{f_1, f_6, f_7, f_8, f_9, f_{11}\}$  and the redundant objective set is given by  $\mathcal{F}_{redn} = \{f_2, f_3, f_4, f_5, f_{10}\}$ .
5. The error incurred when  $\mathcal{F}_{redn}$  is removed from the original objective set (based on Equation 5.3 and Table 5.5a) is given by:

$$\mathcal{E}_{redn} = \sum_{i \in \{2,3,4,5,10\}} \mathcal{E}_i = 0.650058\%. \quad (5.12)$$

The objective reduction results can be validated by the parallel coordinated plots in Figure 5.6, since the solution sets represented in Figures 5.6a and 5.6b are very similar, as the former case corresponds to the original problem formulation while the latter corresponds to the reduced set of objectives  $\mathcal{F}_s$ .

Table 5.4: Car side-impact problem: Approach-A (NL-MVU-PCA based framework), corresponding to  $\mathcal{N}_{\mathcal{NS}}$  (one run)

(a) Correlation matrix ( $R$ )											
	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$f_7$	$f_8$	$f_9$	$f_{10}$	$f_{11}$
$f_1$	1.000	-0.655	0.005	0.045	-0.794	0.052	-0.158	-0.516	-0.781	-0.489	-0.624
$f_2$	-0.655	1.000	-0.354	-0.378	0.787	-0.345	-0.136	0.326	0.857	0.230	0.400
$f_3$	0.005	-0.354	1.000	0.976	-0.336	0.941	0.903	0.673	-0.429	0.747	-0.472
$f_4$	0.045	-0.378	0.976	1.000	-0.373	0.977	0.922	0.648	-0.493	0.718	-0.475
$f_5$	-0.794	0.787	-0.336	-0.373	1.000	-0.353	-0.114	0.278	0.793	0.198	0.517
$f_6$	0.052	-0.345	0.941	0.977	-0.353	1.000	0.964	0.672	-0.527	0.731	-0.377
$f_7$	-0.158	-0.136	0.903	0.922	-0.114	0.964	1.000	0.816	-0.347	0.854	-0.228
$f_8$	-0.516	0.326	0.673	0.648	0.278	0.672	0.816	1.000	0.105	0.990	-0.027
$f_9$	-0.781	0.857	-0.429	-0.493	0.793	-0.527	-0.347	0.105	1.000	0.048	0.520
$f_{10}$	-0.489	0.230	0.747	0.718	0.198	0.731	0.854	0.990	0.048	1.000	-0.059
$f_{11}$	-0.624	0.400	-0.472	-0.475	0.517	-0.377	-0.228	-0.027	0.520	-0.058	1.000

(b) Eigen-decomposition, eigenvalue analysis and selection scores					
	$e_1 = 0.589$	$e_2 = 0.384$	$e_3 = 0.018$	$e_4 = 0.007$	
	$V_1$	$V_2$	$V_3$	$V_4$	$sc_i$
$f_1$	<u>0.770</u>	<u>-0.516</u>	-0.131	-0.181	0.655
$f_2$	0.053	<u>0.192</u>	-0.002	<u>0.160</u>	0.106
$f_3$	0.056	<u>0.166</u>	<u>0.046</u>	<u>0.200</u>	0.099
$f_4$	0.057	<u>0.167</u>	<u>0.049</u>	<u>0.196</u>	0.100
$f_5$	0.058	<u>0.180</u>	<u>0.041</u>	<u>0.178</u>	0.106
$f_6$	<u>-0.329</u>	-0.513	<u>0.339</u>	<u>0.336</u>	0.400
$f_7$	<u>-0.459</u>	-0.415	<u>0.039</u>	-0.310	0.434
$f_8$	<u>-0.271</u>	<u>0.059</u>	<u>-0.795</u>	-0.235	0.199
$f_9$	0.052	<u>0.204</u>	<u>0.024</u>	<u>0.222</u>	0.111
$f_{10}$	0.000	<u>0.149</u>	-0.080	<u>0.138</u>	0.060
$f_{11}$	0.013	<u>0.327</u>	<u>0.471</u>	<u>-0.706</u>	0.147

(c) RCM analysis		
Potential identically correlated set	Eq 4.1(i)	$\hat{\mathcal{S}}_2 = \hat{\mathcal{S}}_5 = \hat{\mathcal{S}}_9 = \{f_2, f_5, f_9\}$ $\hat{\mathcal{S}}_3 = \hat{\mathcal{S}}_4 = \hat{\mathcal{S}}_6 = \{f_3, f_4, f_6\}$ $\hat{\mathcal{S}}_8 = \hat{\mathcal{S}}_{10} = \{f_8, f_{10}\}$
Correlation threshold: $T_{cor}$	Eq 4.1(iii)	$1.0 - 0.588629 \times (1.0 - 2/11) = 0.518395$
Identically correlated set	Eq 4.1(ii)	$\mathcal{S}_2 = \mathcal{S}_5 = \mathcal{S}_9 = \{f_2, f_5, f_9\}$ $\mathcal{S}_3 = \mathcal{S}_4 = \mathcal{S}_6 = \{f_3, f_4, f_6\}$ $\mathcal{S}_8 = \mathcal{S}_{10} = \{f_8, f_{10}\}$



Table 5.5: Car side-impact problem (Continuation from Table 5.4): Approach-A (NL-MVU-PCA based framework), corresponding to  $\mathcal{N}_{\mathcal{NS}}$  (one run)

(a) Error computation					(b) Error based objective-ranking		
Category	With reference to Equation 4.4				Ascending		Cumulative
	$c_i^M$	$max\{\delta_{ij} \cdot R_{ij}\}$	$\mathcal{E}_i$	$\mathcal{E}_i^n$	order		$(\mathcal{E}_i^n)$
$f_1 \in \mathcal{F}_s$	0.451582	–	0.451582	0.479812	$f_{r_1}$	$f_{10}$	0.000089
$f_2 \in \mathcal{F}_{redn}$	0.016044	0.856716 ( $R_{29}$ )	0.002299	0.002443	$f_{r_2}$	$f_4$	0.000411
$f_3 \in \mathcal{F}_{redn}$	0.012768	0.940649 ( $R_{36}$ )	0.000758	0.000805	$f_{r_3}$	$f_3$	0.001216
$f_4 \in \mathcal{F}_{redn}$	0.012984	0.976650 ( $R_{46}$ )	0.000303	0.000322	$f_{r_4}$	$f_2$	0.003659
$f_5 \in \mathcal{F}_{redn}$	0.014777	0.793127 ( $R_{59}$ )	0.003057	0.003248	$f_{r_5}$	$f_5$	0.006907
$f_6 \in \mathcal{F}_s$	0.168046	–	0.168046	0.178551	$f_{r_6}$	$f_9$	0.026000
$f_7 \in \mathcal{F}_s$	0.191671	–	0.191671	0.203653	$f_{r_7}$	$f_{11}$	0.077674
$f_8 \in \mathcal{F}_s$	0.056763	–	0.056763	0.060311	$f_{r_8}$	$f_8$	0.137985
$f_9 \in \mathcal{F}_s$	0.017970	–	0.017970	0.019093	$f_{r_9}$	$f_6$	0.316536
$f_{10} \in \mathcal{F}_{redn}$	0.008763	0.990430 ( $R_{10}$ s)	0.000084	0.000089	$f_{r_{10}}$	$f_7$	0.520189
$f_{11} \in \mathcal{F}_s$	0.048633	–	0.048633	0.051674	$f_{r_{11}}$	$f_1$	1.000000

### Perspective-II: Preference-ranking of all the Objectives

Based on Equation 5.4, the preference-weights for each objective is the same as the normalised errors ( $\mathcal{E}_i^n$ ), shown in Table 5.5a. Sorting of these preference-weights establishes the preference-ranking of all the objectives (Table 5.5b), as follows

$$f_1 \triangleright f_7 \triangleright f_6 \triangleright f_8 \triangleright f_{11} \triangleright f_9 \triangleright f_5 \triangleright f_2 \triangleright f_3 \triangleright f_4 \triangleright f_{10}, \quad (5.13)$$

where  $\triangleright$  denotes *more-important-than* relation.

### Perspective-III: $\delta$ -MOSS Analysis

Based on Equation 5.5 and the sorted (in ascending order)  $\mathcal{E}_i^n$  in Table 5.5b, the  $\delta$ -MOSS analysis is conducted. The  $\delta$ -minimal objective sets for different  $\delta$ 's are as follows:

- $\mathcal{F}_{\{\delta=0.1\}s} = \{f_{r_8}, f_{r_9}, f_{r_{10}}, f_{r_{11}}\} = \{f_8, f_6, f_7, f_1\},$
- $\mathcal{F}_{\{\delta=0.2=0.3\}s} = \{f_{r_9}, f_{r_{10}}, f_{r_{11}}\} = \{f_6, f_7, f_1\},$
- $\mathcal{F}_{\{\delta=0.4=0.5\}s} = \{f_{r_{10}}, f_{r_{11}}\} = \{f_7, f_1\},$

- $\mathcal{F}_{\{\delta=0.6=0.7=0.8=0.9\}s} = \{f_{r_{11}}\} = \{f_1\}$ .

#### Perspective-IV: $k$ -EMOSS Analysis

Based on Equation 5.6 and the sorted (in ascending order)  $\mathcal{E}_i^n$  in Table 5.5b, the  $k$ -EMOSS analysis is conducted. For some different  $k$ 's the  $k$ -minimal objective sets with corresponding  $k$ -minimal errors are the following:

- $p = 10\%$ , i.e.,  $k = \lceil 0.10 \times 11 \rceil = 2$ :  $\mathcal{F}_{\{k=2\}} = \{f_1, f_7\}$  and  $\mathcal{E}_{k=2}^n = 31.65\%$ ,
- $p = 20\%$ , i.e.,  $k = \lceil 0.20 \times 11 \rceil = 3$ :  $\mathcal{F}_{\{k=3\}} = \{f_1, f_6, f_7\}$  and  $\mathcal{E}_{k=3}^n = 13.80\%$ ,
- $p = 30\%$ , i.e.,  $k = \lceil 0.30 \times 11 \rceil = 4$ :  $\mathcal{F}_{\{k=4\}} = \{f_1, f_6, f_7, f_8\}$  and  $\mathcal{E}_{k=4}^n = 7.77\%$ .

### 5.5.2 Sensitivity of the Decision Support to Different Samples of $\mathcal{N}_{\mathcal{NS}}$

This section investigates the sensitivity of the decision support offered by the Approach-A, to different samples of  $\mathcal{N}_{\mathcal{NS}}$  corresponding to the car-side impact problem. For this purpose, the 20 samples of  $\mathcal{N}_{\mathcal{NS}}$  generated by NSGA-II have been used, and the mean and standard deviation for  $|\mathcal{F}_{\{\delta\}}|$  ( $\delta$ -MOSS analysis) and  $\mathcal{E}_k^n$  ( $k$ -EMOSS analysis) are reported in Tables 5.6 and 5.7, respectively. It can be seen that the mean values for  $|\mathcal{F}_{\{\delta\}}|$  and  $\mathcal{E}_k^n$  (in respective cases), strongly comply with those reported in Section 5.5.1 for Perspective-III and Perspective-IV, respectively, given that the standard deviation is not too significant.

The analysis conducted above has been applied to a solution set generated by NSGA-II after 2000 generations. During the optimisation process the preference-weights may vary depending on how much improvement each generation contributes until the final set is achieved. In order to give confidence to the DM that the learnt preference-weights are stable for the problem in consideration, Figure 5.7 shows the preference-weights evolution along the simulation taken every 100 generations. Hence, based on the results obtained it can be inferred that the Approach-A is fairly robust and, the resulting decision support can be relied upon.

Table 5.6: Car side-impact problem:  $\delta$ -MOSS analysis by the Approach-A (NL-MVU-PCA based framework), averaged over 20  $\mathcal{N}_{\mathcal{NS}}$

DM Input	Frequency of occurrence of objectives											Output Indicator
$\delta$ (%)	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$f_7$	$f_8$	$f_9$	$f_{10}$	$f_{11}$	$ \mathcal{F}_{\{\delta\}} $
00	20	07	02	01	01	13	20	20	17	03	20	$6.20 \pm 0.95$
10	20	00	00	00	00	13	20	20	00	00	02	$3.75 \pm 0.55$
20	20	00	00	00	00	13	20	00	00	00	00	$2.65 \pm 0.49$
30	20	00	00	00	00	11	20	00	00	00	00	$2.55 \pm 0.51$
40	20	00	00	00	00	00	20	00	00	00	00	$2.00 \pm 0.00$
50	20	00	00	00	00	00	12	00	00	00	00	$1.60 \pm 0.50$
60	20	00	00	00	00	00	00	00	00	00	00	$1.00 \pm 0.00$
70	20	00	00	00	00	00	00	00	00	00	00	$1.00 \pm 0.00$
80	20	00	00	00	00	00	00	00	00	00	00	$1.00 \pm 0.00$
90	20	00	00	00	00	00	00	00	00	00	00	$1.00 \pm 0.00$

Table 5.7: Car side-impact problem:  $k$ -EMOSS analysis by the Approach-A (NL-MVU-PCA based framework), averaged over 20  $\mathcal{N}_{\mathcal{NS}}$

DM Input	Frequency of occurrence of objectives											Output Indicator
$k$	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$f_7$	$f_8$	$f_9$	$f_{10}$	$f_{11}$	$\mathcal{E}_k^n$
1	20	00	00	00	00	00	00	00	00	00	00	$0.481 \pm 0.051$
2	20	00	00	00	00	00	20	00	00	00	00	$0.260 \pm 0.070$
3	20	00	00	00	00	13	20	07	00	00	00	$0.124 \pm 0.021$
4	20	00	00	00	00	13	20	20	00	00	07	$0.067 \pm 0.021$
5	20	00	00	00	00	13	20	20	07	00	20	$0.029 \pm 0.012$
6	20	04	00	00	00	18	20	20	17	01	20	$0.014 \pm 0.009$
7	20	10	01	01	08	19	20	20	19	02	20	$0.007 \pm 0.005$
8	20	19	02	01	16	19	20	20	20	03	20	$0.003 \pm 0.003$
9	20	20	15	03	19	20	20	20	20	03	20	$0.001 \pm 0.001$
10	20	20	20	17	20	20	20	20	20	03	20	$0.000 \pm 0.000$

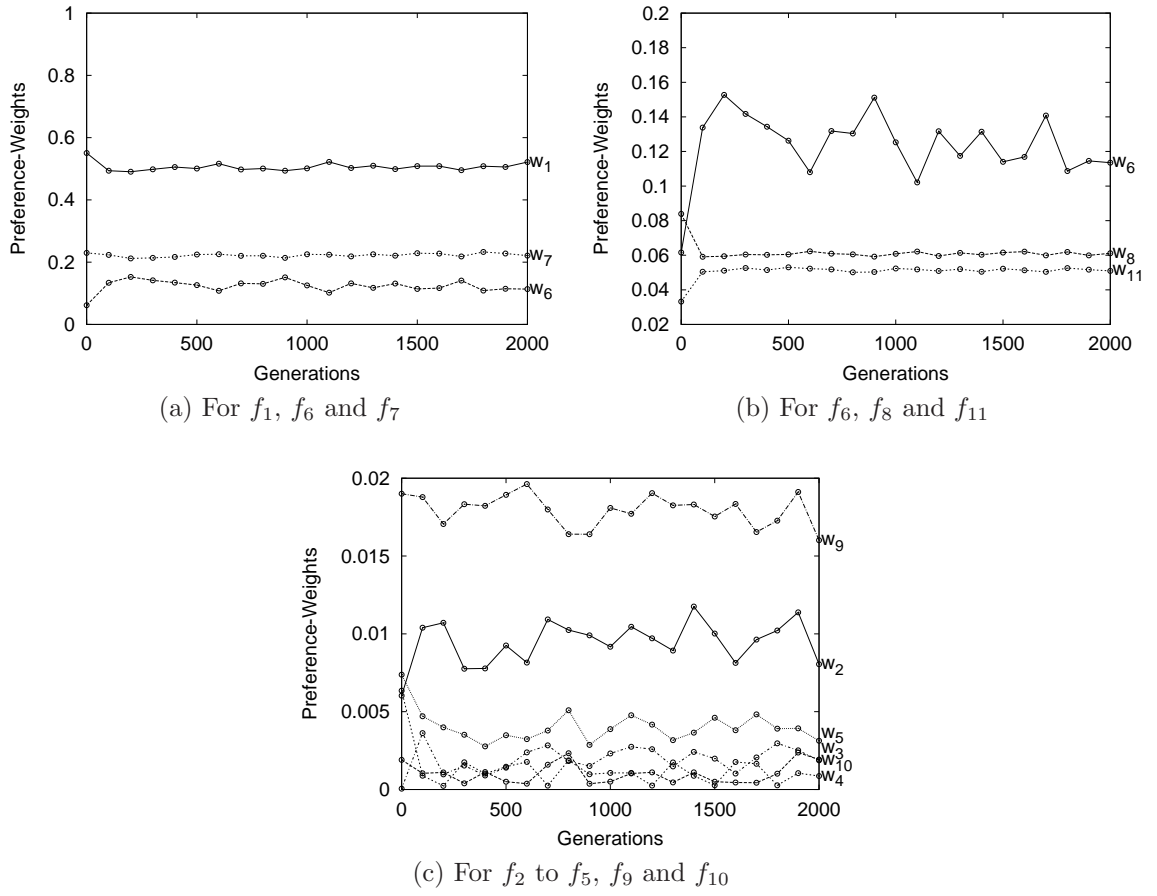


Figure 5.7: Car side-impact problem: Evolution of objective preference-weights captured by the Approach-A (NL-MVU-PCA based framework), averaged over 20  $\mathcal{N}_{\mathcal{S}}$ . The weights are split in subfigures due to varying scales.

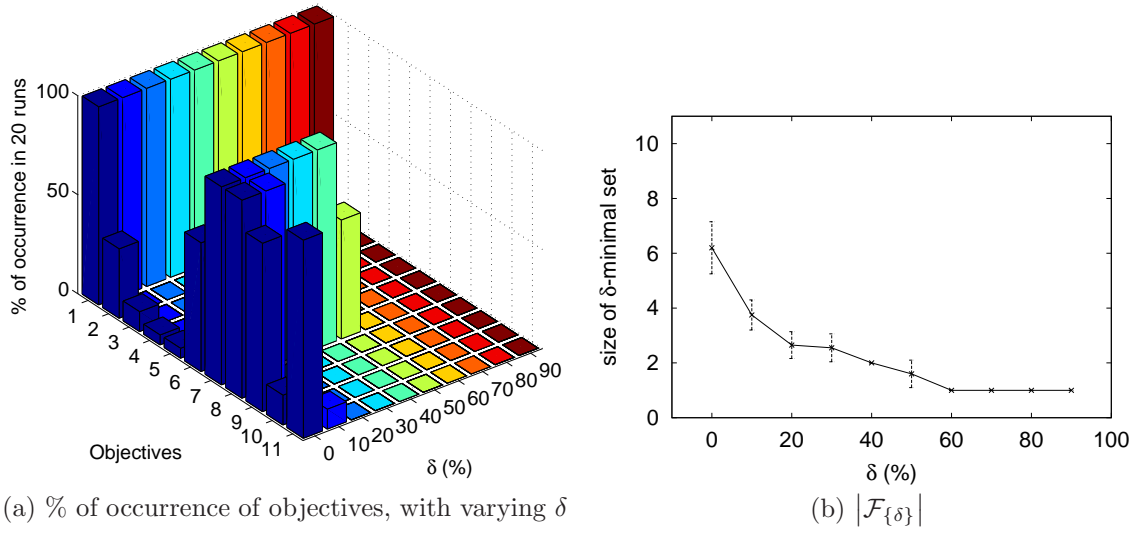


Figure 5.8: Car side-impact problem: Visual representation of the  $\delta$ -MOSS analysis, by the Approach-A (NL-MVU-PCA based framework), averaged over 20  $\mathcal{N}_{NS}$ .

### 5.5.3 Visual Representation of the Results

The DM may wish to be presented with simple and straightforward information that would help him/her to take decisions in a systematic way when confronted with a multi-objective problem. In this way it is possible to use an MCDM based MOEA to generate a set of solutions representative of the problem objective characteristics which are then revealed by further analysis. While Figures 5.8b and 5.9b gives a general idea of the problem redundancy, Figures 5.8a and 5.9a inform the DM about the objective set of a specified size that yield minimal error.

### 5.5.4 Key Highlights of the Decision Support Offered by the Proposed Framework

The key highlights of the proposed decision support offered by the Approach-A, are as follows:

1. The revelation of an *essential* objective set (Perspective-I) informs the DM that an objective set  $\mathcal{F}_s = \{f_1, f_6, f_7, f_8, f_9, f_{11}\}$  incurs only in an error of 0.650058% and therefore the remaining 5 objectives can be discarded which facilitates the problem

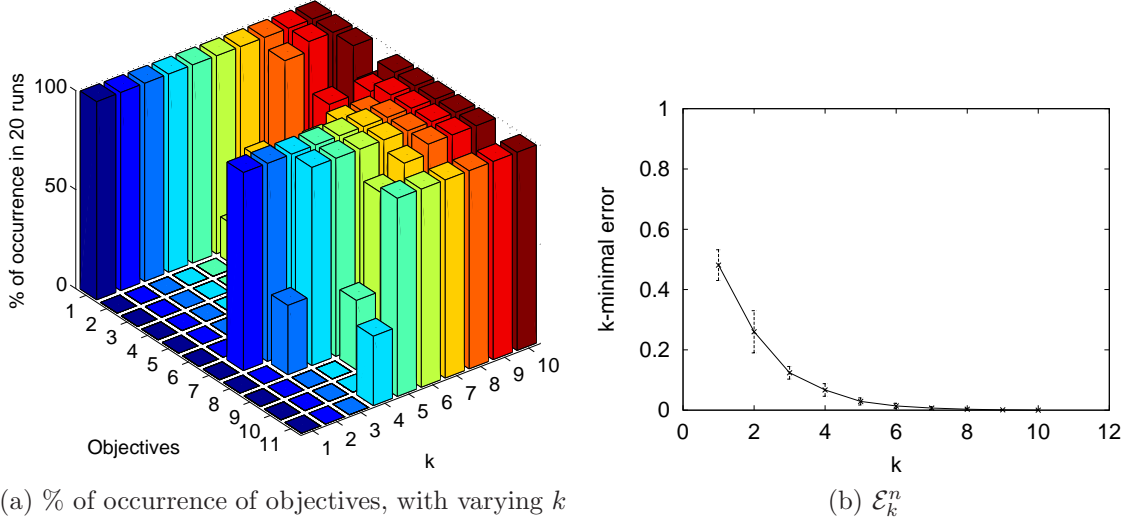


Figure 5.9: Car side-impact problem: Visual representation of the  $k$ -EMOSS analysis, by the Approach-A (NL-MVU-PCA based framework), averaged over 20  $\mathcal{N}_{NS}$ .

analysis since the DM only needs to handle 6 objectives instead of 11. The abdomen load ( $f_2$ ), viscous criterion ( $f_3, f_4, f_5$ ) and velocity of V-pillar at middle point ( $f_{10}$ ) can be ignored.

2. The preference-ranking of all the objectives (Perspective-II) reveals  $f_1 \triangleright f_7 \triangleright f_6 \triangleright f_8 \triangleright f_{11} \triangleright f_9 \triangleright f_5 \triangleright f_2 \triangleright f_3 \triangleright f_4 \triangleright f_{10}$ , where  $\triangleright$  denotes *more-important-than* relation, and assists in arriving at the following interpretation:

- Since  $f_{11} \triangleright f_9$ , minimising the front door at V-pillar ( $f_{11}$ ) is more important than minimising the public force ( $f_9$ ), and, more important than those two is minimising the rib deflections ( $f_6$ - $f_8$ ) and, fundamentally, the weight ( $f_1$ ).
- Since  $f_7 \triangleright f_6 \triangleright f_8$ , the most important rib deflection is the middle and the lower is the least important.

3. In a case where the DM is able to allow an error of at least 30% then he/she only needs to deal with minimisation of objective  $f_4$  which considerably simplifies the optimisation process. However, supposing that an error of 20% is considered safer to the decision-making process then only  $f_3$  and  $f_4$  need to be considered.

4. In a case where the DM is able to allow an error of at least 31.6536%, accounting for only two objectives, namely,  $f_1$  and  $f_7$  will suffice. On the other hand, if the DM wants to account for at least three objectives which correspond to minimum error, then  $f_1$ ,  $f_6$  and  $f_7$  need to be accounted for, corresponding to 13.7985% of error.

## 5.6 Experimental Results on Storm Drainage System Problem

This is a five-objective problem with seven constraints that relates to the optimisation of a storm drainage system in an urban area. A detailed description of the problem can be found in [160] and the physical meaning of the objectives and constraints is provided by Equation 4.14 (Chapter 4).

### 5.6.1 Demonstration of the Approach-A (NL-MVU-PCA based Framework)

The analysis that follows corresponds to the application of the Approach-A to one of the 20 non-dominated solution sets.

#### Perspective-I: Revelation of an *Essential* Objective Set

The application of Approach-A is detailed in Table 5.8. The steps are:

1. The correlation matrix ( $R$ ) and the eigenvalues/eigenvectors of matrix  $K$  are shown in Tables 5.8a and 5.8b, respectively.
2. The eigenvalues analysis shown in Table 5.8b reveals that: (i) the number of principal components ( $N_v$ ) needed to meet the variance threshold ( $\theta = 0.997$ ) is two, and (ii) in those two principal components all five objectives are selected, leading to  $\mathcal{F}_e = \mathcal{F}_0$ .

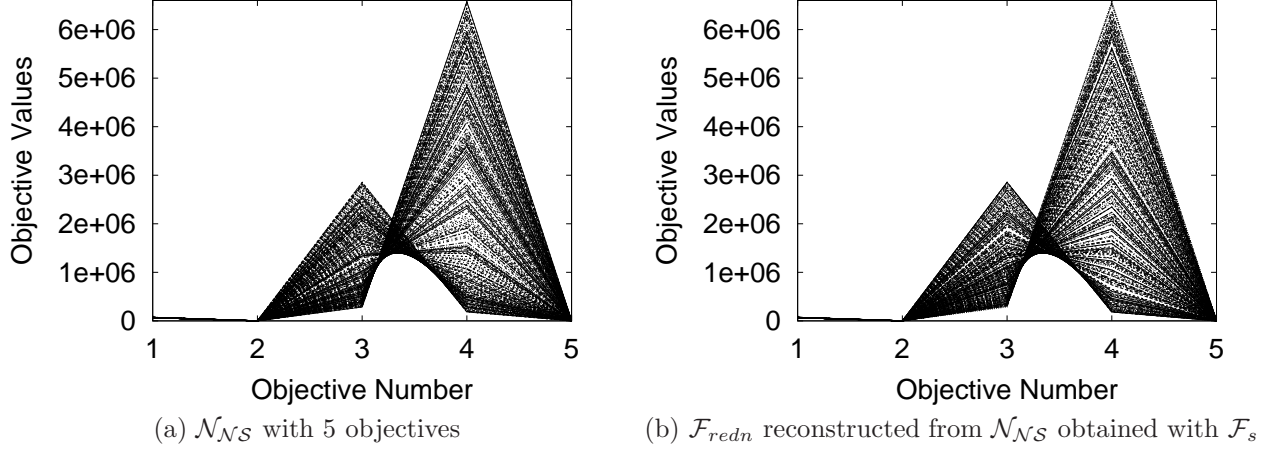


Figure 5.10: Storm drainage system problem: Parallel coordinate plots, corresponding to one run of NSGA-II.

3. The RCM analysis shown in Table 5.8c reveals two *potentially identically* correlated objective subsets, with only one distinct set of composition— $\{f_1, f_3\}$ . Due to a relative low  $T_{cor} = 0.210$  indicative of redundancy in the problem, the two objective subsets are considered *identically* correlated since  $R_{13} = 0.998$ .
4. The selection score ( $sc_i$ ) computed in Table 5.8b selects  $f_3$  as the most significant objective from subsets  $\mathcal{S}_1$  and  $\mathcal{S}_3$  which leads to  $\mathcal{F}_{redn} = \{f_1\}$  and  $\mathcal{F}_s = \{f_2, f_3, f_4, f_5\}$ .
5. The error incurred when  $\mathcal{F}_{redn}$  is removed from the original objective set (based on Equation 5.3 and Table 5.8d) is given by

$$\mathcal{E}_{redn} = \sum_{i \in \{1\}} \mathcal{E}_i = 0.002841\%. \quad (5.14)$$

The objective reduction results can be validated by the parallel coordinated plots in Figure 5.10, since the solution sets represented in Figures 5.10a and 5.10b are very identical. The former case corresponds to the original problem formulation while the latter corresponds to the reduced set of objectives  $\mathcal{F}_s$ .



Table 5.8: Storm drainage system problem: Approach-A based on  $\mathcal{N}_{\mathcal{S}}$  (one run)

(a) Correlation matrix ( $R$ )						(b) Eigen-decomposition, eigenvalue analysis and selection scores			
	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$e_1= 0.988$ $e_2= 0.012$			
						$V_1$	$V_2$	$sc_i$	
$f_1$	1.000	-0.245	0.998	-0.920	-0.516	$f_1$	<u>0.126</u>	-0.341	0.129
$f_2$	-0.245	1.000	-0.234	0.333	-0.475	$f_2$	<u>0.125</u>	-0.344	0.127
$f_3$	0.998	-0.234	1.000	-0.921	-0.527	$f_3$	<u>0.477</u>	<u>0.757</u>	0.480
$f_4$	-0.920	0.333	-0.921	1.000	0.571	$f_4$	-0.852	0.272	0.845
$f_5$	-0.516	-0.475	-0.527	0.571	1.000	$f_5$	<u>0.124</u>	-0.344	0.127

(c) RCM analysis		
Potential identically correlated set	Eq 4.1(i)	$\hat{\mathcal{S}}_1 = \hat{\mathcal{S}}_3 = \{f_1, f_3\}$
Correlation threshold: $T_{cor}$	Eq 4.1(iii)	$1.0 - 0.988121 \times 1/5 = 0.20950$
Identically correlated set	Eq 4.1(ii)	$\mathcal{S}_1 = \mathcal{S}_3 = \{f_1, f_3\}$

(d) Error computation					(e) Error based objective-ranking		
Category	With reference to Equation 4.4				Ascending order		Cumulative
	$c_i^M$	$max\{\delta_{i\ j} \cdot R_{ij}\}$	$\mathcal{E}_i$	$\mathcal{E}_i^n$			$\mathcal{E}_i^n$
$f_1 \in \mathcal{F}_{redn}$	0.017119	0.998340 ( $R_{13}$ )	0.000028	0.000029	$f_{r_1}$	$f_1$	0.000029
$f_2 \in \mathcal{F}_s$	0.016812	—	0.016812	0.017105	$f_{r_2}$	$f_5$	0.016929
$f_3 \in \mathcal{F}_s$	0.231440	—	0.231440	0.235464	$f_{r_3}$	$f_2$	0.034034
$f_4 \in \mathcal{F}_s$	0.718017	—	0.718017	0.730502	$f_{r_4}$	$f_3$	0.269498
$f_5 \in \mathcal{F}_s$	0.016612	—	0.016612	0.016900	$f_{r_5}$	$f_4$	1.000000

### Perspective-II: Preference-ranking of all the Objectives

Based on Equation 5.4, the preference-weights for each objective is the same as the normalised errors ( $\mathcal{E}_i^n$ ), shown in Table 5.8d. Sorting of these preference-weights establishes the preference-ranking of all the objectives (Table 5.8e), as follows

$$f_4 \triangleright f_3 \triangleright f_2 \triangleright f_5 \triangleright f_1, \quad (5.15)$$

where  $\triangleright$  denotes *more-important-than* relation.

### Perspective-III: $\delta$ -MOSS Analysis

Based on Equation 5.5 and the sorted (in ascending order)  $\mathcal{E}_i^n$  in Table 5.8e, the  $\delta$ -MOSS analysis is conducted. The  $\delta$ -minimal objective sets for different  $\delta$ 's are as follows:

- $\mathcal{F}_{\{\delta=0.1=0.2\}s} = \{f_3, f_4\},$
- $\mathcal{F}_{\{\delta>0.2695\}s} = \{f_4\}.$

### Perspective-IV: $k$ -EMOSS Analysis

Based on Equation 5.6 and the sorted (in ascending order)  $\mathcal{E}_i^n$  in Table 5.8e, the  $k$ -EMOSS analysis is conducted. For some different  $k$ 's the  $k$ -minimal objective sets with corresponding  $k$ -minimal errors are the following:

- $p = 10\%$ , i.e.,  $k = \lceil 0.10 \times 5 \rceil = 1$ :  $\mathcal{F}_{\{k=1\}} = \{f_4\}$  and  $\mathcal{E}_{k=1}^n = 26.95\%$ ,
- $p = 30\%$ . i.e.,  $k = \lceil 0.30 \times 5 \rceil = 2$ :  $\mathcal{F}_{\{k=2\}} = \{f_3, f_4\}$  and  $\mathcal{E}_{k=2}^n = 03.40\%$ ,
- $p = 50\%$ , i.e.,  $k = \lceil 0.50 \times 5 \rceil = 3$ :  $\mathcal{F}_{\{k=3\}} = \{f_2, f_3, f_4\}$  and  $\mathcal{E}_{k=3}^n = 01.69\%$ .

Table 5.9: Storm drainage system problem:  $\delta$ -MOSS analysis by the Approach-A (NL-MVU-PCA based framework), averaged over 20  $\mathcal{N}_{\mathcal{NS}}$ 

DM Input	Frequency of occurrence of objectives					Output Indicator
$\delta$ (%)	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$ \mathcal{F}_{\{\delta\}} $
00	00	20	20	20	20	$4.00 \pm 0.00$
10	00	00	20	20	00	$2.00 \pm 0.00$
20	00	00	20	20	00	$2.00 \pm 0.00$
30	00	00	00	20	00	$1.00 \pm 0.00$
40	00	00	00	20	00	$1.00 \pm 0.00$
50	00	00	00	20	00	$1.00 \pm 0.00$
60	00	00	00	20	00	$1.00 \pm 0.00$
70	00	00	00	20	00	$1.00 \pm 0.00$
80	00	00	00	20	00	$1.00 \pm 0.00$
90	00	00	00	20	00	$1.00 \pm 0.00$

 Table 5.10: Storm drainage system problem:  $k$ -EMOSS analysis by the Approach-A (NL-MVU-PCA based framework), averaged over 20  $\mathcal{N}_{\mathcal{NS}}$ 

DM Input	Frequency of occurrence of objectives					Output Indicator
$k$	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$\mathcal{E}_k^n$
1	00	00	00	20	00	$0.274 \pm 0.003$
2	00	00	20	20	00	$0.033 \pm 0.001$
3	00	20	20	20	00	$0.016 \pm 0.000$
4	00	20	20	20	20	$0.000 \pm 0.000$

### 5.6.2 Sensitivity of the Decision Support to Different Samples of $\mathcal{N}_{\mathcal{NS}}$

This section investigates the sensitivity of the decision support offered by the Approach-A, to different samples of  $\mathcal{N}_{\mathcal{NS}}$  corresponding to the storm drainage system problem. For this purpose, the 20 samples of  $\mathcal{N}_{\mathcal{NS}}$  generated by NSGA-II have been used, and the mean and standard deviation for  $|\mathcal{F}_{\{\delta\}}|$  ( $\delta$ -MOSS analysis) and  $\mathcal{E}_k^n$  ( $k$ -EMOSS analysis) are reported in Tables 5.9 and 5.10. It can be seen that the mean values for  $|\mathcal{F}_{\{\delta\}}|$  and  $\mathcal{E}_k^n$  (in respective cases), strongly comply with those reported in Section 5.6.1 for Perspective-III and Perspective-IV, respectively, given that the standard deviation is relatively low.

To give confidence to the DM that the learnt preference-weights are stable for the problem under consideration, the preference-weights evolution along the simulation is shown in

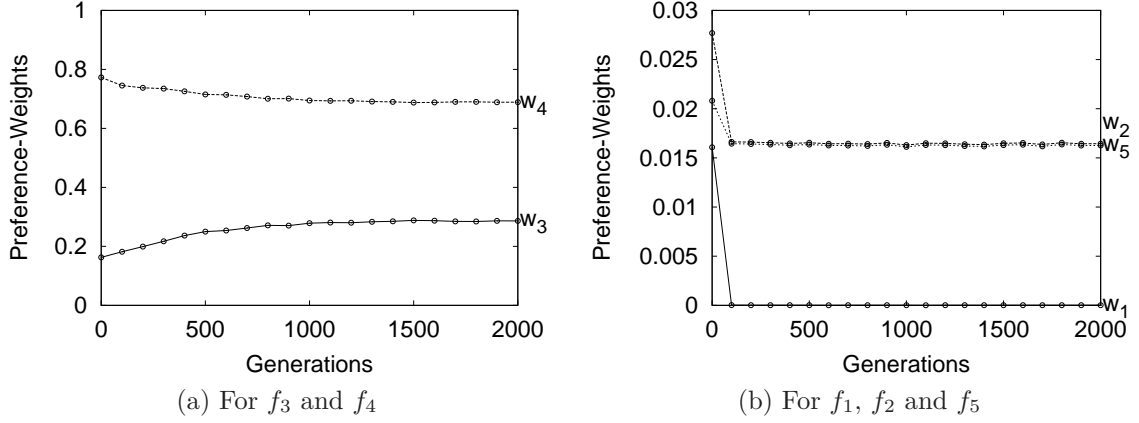


Figure 5.11: Storm drainage system problem: Evolution of objective preference-weights captured by the Approach-A (NL-MVU-PCA based framework), averaged over 20  $\mathcal{N}_{NS}$ . The weights are split in subfigures due to varying scales.

Figure 5.11. The weights are averaged out of 20 runs and taken every 100 generations. Hence, based on the results obtained, it can be inferred that the Approach-A is fairly robust and the resulting decision support can be relied upon.

### 5.6.3 Visual Representation of the Results

The information shared with the DM can be shown in an uncomplicated and time-efficient manner as represented in Figures 5.12 and 5.13, for  $\delta$ -MOSS and  $k$ -EMOSS analysis, respectively. While the former case gives a general idea of the problem redundancy, the latter informs the DM about the objective set of a specified size that yields a minimal error.

### 5.6.4 Key Highlights of the Decision Support Offered by the Proposed Framework

The key highlights of the proposed decision support offered by the Approach-A, are as follows:

- The revelation of an *essential* objective set (Perspective-I) informs the DM that an objective set  $\mathcal{F}_s = \{f_2, f_3, f_4, f_5\}$  incurs an error of 0.002841% and therefore objective  $f_1$  can be discarded. Minimising the drainage network cost ( $f_1$ ) is highly correlated

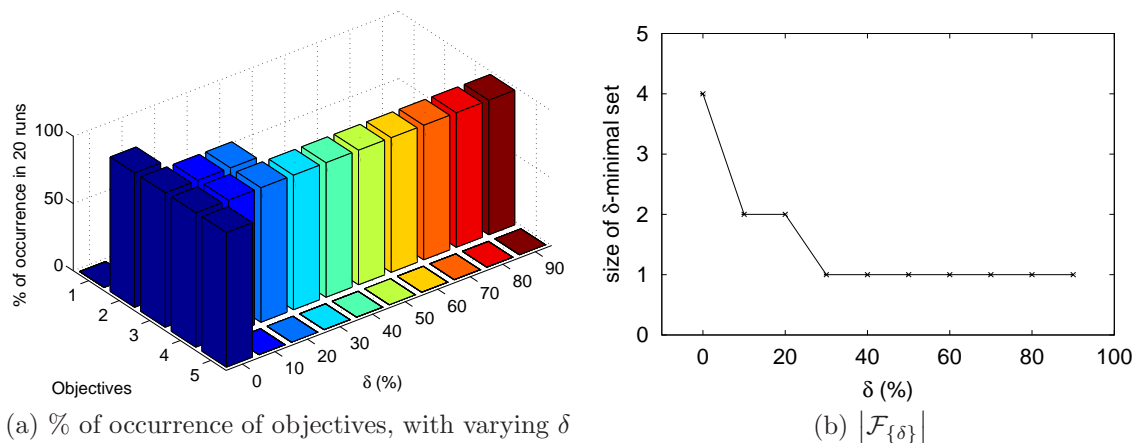


Figure 5.12: Storm drainage system problem: Visual representation of the  $\delta$ -MOSS analysis, by the Approach-A (NL-MVU-PCA based framework), averaged over 20  $\mathcal{N}_{\mathcal{NS}}$ .

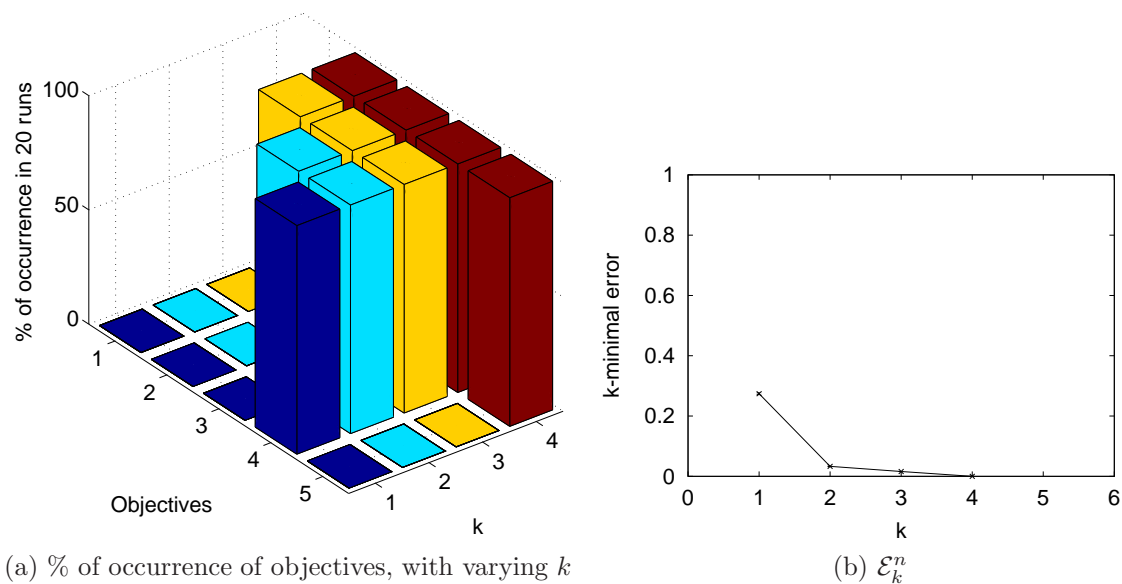


Figure 5.13: Storm drainage system problem: Visual representation of the  $k$ -EMOSS analysis, by the Approach-A (NL-MVU-PCA based framework), averaged over 20  $\mathcal{N}_{\mathcal{NS}}$ .

with minimisation of the treatment facility cost ( $f_3$ ) due to high correlation (Table 5.8a  $R_{13} = 0.998$ ) and since  $sc_3 > sc_1$ , i.e.  $0.480 > 0.129$ , hence, the drainage network cost can be ignored.

- The preference-ranking of all the objectives (Perspective-II) reveals  $f_4 \triangleright f_3 \triangleright f_2 \triangleright f_5 \triangleright f_1$ , where  $\triangleright$  denotes *more-important-than* relation, and assists with arriving at the following interpretation:
  - It is very important to prevent flood damage since the most important objective is the expected flood damage cost ( $f_4$ ).
  - Since  $f_3 \triangleright f_2 \triangleright f_5$ , the cost associated with the facility treatment is more important than with facility storage and the expected economic loss due to flood is the least important objective.
- In a case where the DM is able to allow an error of at least 26.9498% then he/she only needs to deal with minimisation of objective  $f_4$  which considerably simplifies the optimisation process. However, if an error of 3.4034% is considered safer to the decision-making process, only  $f_3$  and  $f_4$  need to be considered.

## 5.7 Experimental Results on Work Roll Cooling Design Problem

This problem deals with the design optimisation of a work roll based on the cooling performance. The hot metal rolling process is a mechanical technique used by industry for shaping metal bars. The problem consists of six objectives and seven variables. While the detailed description of the problem can be found in [165] the physical meaning of the objectives are

given by:

$$\left. \begin{aligned}
 &\text{Minimise:} \\
 &(x_1, \dots, x_7) \\
 &f_1(\mathbf{x}) \equiv \text{Change in temperature at roll surface,} \\
 &f_2(\mathbf{x}) \equiv \text{Radial stress at the roll surface,} \\
 &f_3(\mathbf{x}) \equiv \text{Change in temperature at 9mm depth,} \\
 &f_4(\mathbf{x}) \equiv \text{Radial stress at 9mm depth,} \\
 &f_5(\mathbf{x}) \equiv \text{Change in temperature at 15mm depth,} \\
 &f_6(\mathbf{x}) \equiv \text{Radial stress at 15mm depth,}
 \end{aligned} \right\} \quad (5.16)$$

and the 7 variables physical meaning are given by:

$$\left. \begin{aligned}
 &x_1 \equiv \text{Roll / Stock contact HTC,} \\
 &x_2 \equiv \text{Stock temperature,} \\
 &x_3 \equiv \text{Roll / Stock Contact length,} \\
 &x_4 \equiv \text{Cooling HTC,} \\
 &x_5 \equiv \text{Roll speed,} \\
 &x_6 \equiv \text{Roll temperature,} \\
 &x_7 \equiv \text{Delay time.}
 \end{aligned} \right\} \quad (5.17)$$

The objectives  $\{f_1, f_3, f_5\}$  are related to the rolling temperature and it is expected for them not to conflict. The same is expected for objectives  $\{f_2, f_4, f_6\}$  since they are related to the radial stress, and it is anticipated that they will conflict with the objectives related to the rolling temperature.

### 5.7.1 Demonstration of the Approach-A (NL-MVU-PCA based Framework)

The analysis that follows corresponds to the application of the Approach-A to one of the 20 non-dominated solution sets.

### Perspective-I: Revelation of an *Essential* Objective Set

The application of Approach-A is detailed in Table 5.11 and the steps are:

1. The correlation matrix ( $R$ ) and the eigenvalues/eigenvectors of matrix  $K$  are shown in Tables 5.11a and 5.11b, respectively.
2. The eigenvalues analysis shown in Table 5.11b reveals that: (i) the number of principal components ( $N_v$ ) needed to meet the variance threshold ( $\theta = 0.997$ ) is two, and (ii) in those two principal components all five objectives are selected, leading to  $\mathcal{F}_e = \mathcal{F}_0$ .
3. The RCM analysis shown in Table 5.11c reveals: (i) six *potentially identically* correlated objective subsets, with only two distinct set of composition— $\{f_1, f_3, f_5\}$  comprising all the temperature related objectives, and  $\{f_2, f_4, f_6\}$  comprising all the stress related objectives, (ii) a relative low  $T_{cor} = 0.177$  indicative of redundancy in the problem, and (iii) six *identically* correlated objective subsets, with two distinct sets of composition. The captured correlations are considered very high comparatively to  $T_{cor}$  since  $R_{24} = 0.975$ ,  $R_{31} = 0.994$ ,  $R_{51} = 0.989$ , and  $R_{64} = 0.998$ , leading to  $\mathcal{S}_1 = \hat{\mathcal{S}}_1$ ,  $\mathcal{S}_2 = \hat{\mathcal{S}}_2$ ,  $\mathcal{S}_3 = \hat{\mathcal{S}}_3$ ,  $\mathcal{S}_4 = \hat{\mathcal{S}}_4$ ,  $\mathcal{S}_5 = \hat{\mathcal{S}}_5$ , and  $\mathcal{S}_6 = \hat{\mathcal{S}}_6$ .
4. Based on the selection score ( $sc_i$ ) computed in Table 5.11b, the picked objectives as essential are as follows: (i)  $f_1$  from subsets  $\mathcal{S}_1$ ,  $\mathcal{S}_3$ , and  $\mathcal{S}_5$ ; (ii)  $f_4$  from subsets  $\mathcal{S}_2$ ,  $\mathcal{S}_4$ , and  $\mathcal{S}_6$ . As a results,  $\mathcal{F}_{redn} = \{f_2, f_3, f_5\}$  and  $\mathcal{F}_s = \{f_1, f_4\}$ .
5. The error incurred when  $\mathcal{F}_{redn}$  is removed from the original objective set (based on Equation 5.3 and Table 5.11d) is given by

$$\mathcal{E}_{redn} = \sum_{i \in \{2,3,5,6\}} \mathcal{E}_i = 0.612534\%. \quad (5.18)$$

The objective reduction results can be validated by the parallel coordinated plots in Figure 5.14, since the solution sets represented in Figures 5.14a and 5.14b are very similar.



Table 5.11: Work roll cooling design problem: Approach-A based on  $\mathcal{N}_{\mathcal{NS}}$  (one run)

(a) Correlation matrix ( $R$ )							(b) Eigen-decomposition, eigenvalue analysis and selection scores			
	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$e_1 = 0.986$ $e_2 = 0.013$			
							$V_1$	$V_2$	$sc_i$	
$f_1$	1.000	-0.964	0.994	-0.995	0.989	-0.996	$f_1$	<u>-0.458</u>	<u>-0.079</u>	0.452
$f_2$	-0.964	1.000	-0.955	0.975	-0.943	0.960	$f_2$	0.357	<u>0.815</u>	0.363
$f_3$	0.994	-0.955	1.000	-0.996	0.999	-0.999	$f_3$	<u>-0.398</u>	<u>-0.009</u>	0.393
$f_4$	-0.995	0.975	-0.996	1.000	-0.991	0.998	$f_4$	<u>0.466</u>	<u>-0.241</u>	0.463
$f_5$	0.989	-0.943	0.998	-0.991	1.000	-0.997	$f_5$	<u>-0.364</u>	0.035	0.359
$f_6$	-0.996	0.960	-0.999	0.998	-0.997	1.000	$f_6$	0.395	<u>-0.520</u>	0.396

(c) RCM analysis			
Potential identically correlated set	Eq 4.1(i)	$\hat{\mathcal{S}}_1 = \hat{\mathcal{S}}_3 = \hat{\mathcal{S}}_5 = \{f_1, f_3, f_5\}$	
		$\hat{\mathcal{S}}_2 = \hat{\mathcal{S}}_4 = \hat{\mathcal{S}}_6 = \{f_2, f_4, f_6\}$	
Correlation threshold: $T_{cor}$	Eq 4.1(iii)	$1.0 - 0.986437 \times 1/6 = 0.177969$	
Identically correlated set	Eq 4.1(ii)	$\mathcal{S}_1 = \mathcal{S}_3 = \mathcal{S}_5 = \{f_1, f_3, f_5\}$	
		$\mathcal{S}_2 = \mathcal{S}_4 = \mathcal{S}_6 = \{f_2, f_4, f_6\}$	

(d) Error computation					(e) Error based objective-ranking		
Category	With reference to Equation 4.4				Ascending		Cumulative
	$c_i^M$	$max\{\delta_{ij} \cdot R_{ij}\}$	$\mathcal{E}_i$	$\mathcal{E}_i^n$	order		$\mathcal{E}_i^n$
$f_1 \in \mathcal{F}_s$	0.206287	—	0.206287	0.482183	$f_{r_1}$	$f_6$	0.000762
$f_2 \in \mathcal{F}_{redn}$	0.134376	0.974733 ( $R_{24}$ )	0.003395	0.007936	$f_{r_2}$	$f_3$	0.003098
$f_3 \in \mathcal{F}_{redn}$	0.156181	0.993600 ( $R_{31}$ )	0.001000	0.002336	$f_{r_3}$	$f_5$	0.006381
$f_4 \in \mathcal{F}_s$	0.215406	—	0.215406	0.503499	$f_{r_4}$	$f_2$	0.014318
$f_5 \in \mathcal{F}_{redn}$	0.130606	0.989246 ( $R_{51}$ )	0.001405	0.003283	$f_{r_5}$	$f_1$	0.496501
$f_6 \in \mathcal{F}_{redn}$	0.157144	0.997926 ( $R_{64}$ )	0.000326	0.000762	$f_{r_6}$	$f_4$	1.000000

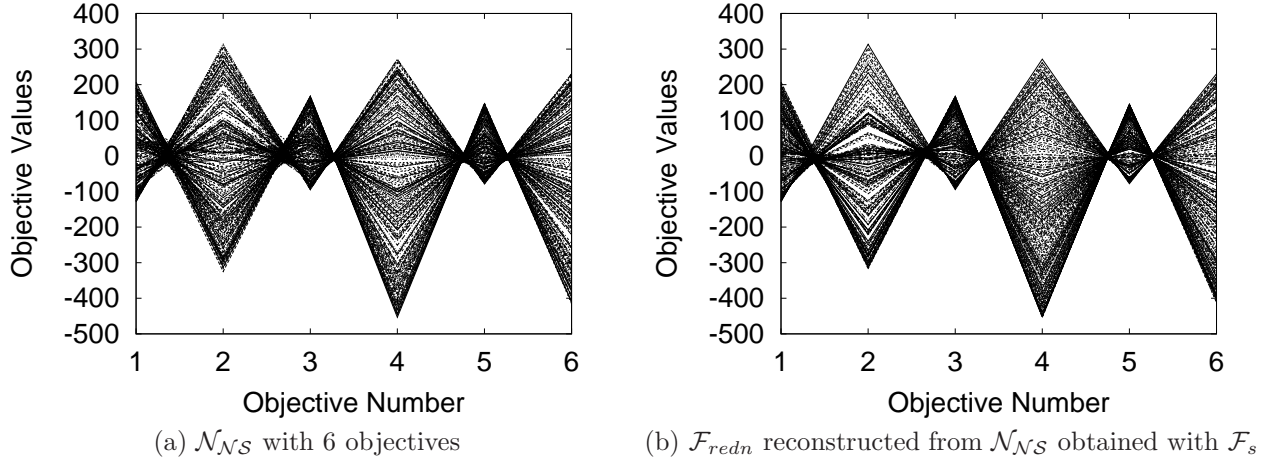


Figure 5.14: Work roll cooling design problem: Parallel coordinate plots (normalised), corresponding to one run of NSGA-II.

The former case corresponds to the original problem formulation and the latter corresponds to the reduced set of objectives  $\mathcal{F}_s$ .

### Perspective-II: Preference-ranking of all the Objectives

Based on Equation 5.4, the preference-weights for each objective is the same as the normalised errors ( $\mathcal{E}_i^n$ ), shown in Table 5.11d. Sorting of these preference-weights establishes the preference-ranking of all the objectives (Table 5.11e), as follows

$$f_4 \triangleright f_1 \triangleright f_2 \triangleright f_5 \triangleright f_3 \triangleright f_6, \quad (5.19)$$

where  $\triangleright$  denotes *more-important-than* relation.

### Perspective-III: $\delta$ -MOSS Analysis

Based on Equation 5.5 and the sorted (in ascending order)  $\mathcal{E}_i^n$  in Table 5.11e, the  $\delta$ -MOSS analysis is conducted. The  $\delta$ -minimal objective sets for different  $\delta$ 's are as follows:

- $\mathcal{F}_{\{\delta=0.1=0.2=0.3=0.4\}s} = \{f_1, f_4\},$
- $\mathcal{F}_{\{\delta>0.4965\}s} = \{f_4\}.$

### Perspective-IV: $k$ -EMOSS Analysis

Based on Equation 5.6 and the sorted (in ascending order)  $\mathcal{E}_i^n$  in Table 5.11e, the  $k$ -EMOSS analysis is conducted. For some different  $k$ 's the  $k$ -minimal objective sets with corresponding  $k$ -minimal errors are the following:

- $p = 10\%$ , i.e.,  $k = \lceil 0.10 \times 6 \rceil = 1$ :  $\mathcal{F}_{\{k=1\}} = \{f_4\}$  and  $\mathcal{E}_{k=1}^n = 49.65\%$ ,
- $p = 20\%$ , i.e.,  $k = \lceil 0.20 \times 6 \rceil = 2$ :  $\mathcal{F}_{\{k=2\}} = \{f_1, f_4\}$  and  $\mathcal{E}_{k=2}^n = 01.43\%$ ,
- $p = 40\%$ , i.e.,  $k = \lceil 0.40 \times 6 \rceil = 3$ :  $\mathcal{F}_{\{k=3\}} = \{f_1, f_2, f_4\}$  and  $\mathcal{E}_{k=3}^n = 00.63\%$ .

### 5.7.2 Sensitivity of the Decision Support to Different Samples of $\mathcal{N}_{\mathcal{NS}}$

This section investigates the sensitivity of the decision support offered by the Approach-A, to different samples of  $\mathcal{N}_{\mathcal{NS}}$  corresponding to the work roll cooling design problem. NSGA-II generated 20 samples of  $\mathcal{N}_{\mathcal{NS}}$  and their mean and standard deviation for  $|\mathcal{F}_{\{\delta\}}|$  ( $\delta$ -MOSS analysis) and  $\mathcal{E}_k^n$  ( $k$ -EMOSS analysis) are reported in Tables 5.12 and 5.13, respectively. It can be seen that the mean values for  $|\mathcal{F}_{\{\delta\}}|$  and  $\mathcal{E}_k^n$  (in respective cases), strongly comply with those reported in Section 5.7.1 for Perspective-III and Perspective-IV, respectively, given that the standard deviation is relatively low.

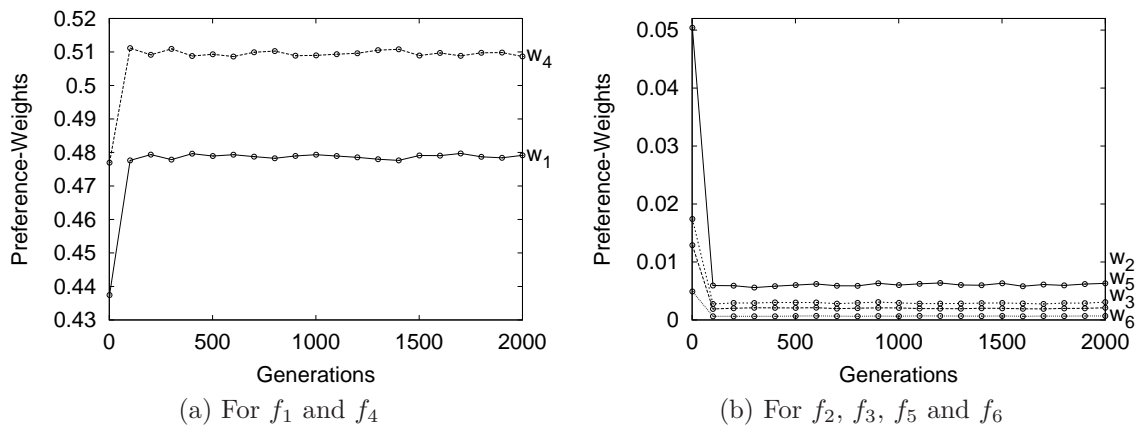
To give confidence to the DM that the learnt preference-weights are stable for the problem in consideration, the preference-weights evolution along the simulation is shown in Figure 5.15. The weights are averaged out of 20 runs and taken every 100 generations. Hence, based on the results obtained it can be inferred that Approach-A is fairly robust and, the resulting decision support can be relied upon.

Table 5.12: Work roll cooling design problem:  $\delta$ -MOSS analysis by the Approach-A (NL-MVU-PCA based framework), averaged over 20  $\mathcal{N}_{\mathcal{NS}}$ 

DM Input	Frequency of occurrence of objectives						Output Indicator
$\delta$ (%)	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$ \mathcal{F}_{\{\delta\}} $
00	20	00	00	20	00	00	$2.00 \pm 0.00$
01	20	00	00	20	00	00	$2.00 \pm 0.00$
02	20	00	00	20	00	00	$2.00 \pm 0.00$
03	20	00	00	20	00	00	$2.00 \pm 0.00$
04	20	00	00	20	00	00	$2.00 \pm 0.00$
05	00	00	00	20	00	00	$1.00 \pm 0.00$
06	00	00	00	20	00	00	$1.00 \pm 0.00$
07	00	00	00	20	00	00	$1.00 \pm 0.00$
08	00	00	00	20	00	00	$1.00 \pm 0.00$
09	00	00	00	20	00	00	$1.00 \pm 0.00$

Table 5.13: Work roll cooling design problem:  $k$ -EMOSS analysis by the Approach-A (NL-MVU-PCA based framework), averaged over 20  $\mathcal{N}_{\mathcal{NS}}$ 

DM Input	Frequency of occurrence of objectives						Output Indicator
$k$	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$\mathcal{E}_k^n$
1	00	00	00	20	00	00	$0.491 \pm 0.003$
2	20	00	00	20	00	00	$0.013 \pm 0.001$
3	20	20	00	20	00	00	$0.006 \pm 0.001$
4	20	20	00	20	20	00	$0.003 \pm 0.000$
5	20	20	20	20	20	00	$0.001 \pm 0.000$

Figure 5.15: Work roll cooling design problem: Evolution of objective preference-weights captured by the Approach-A (NL-MVU-PCA based framework), averaged over 20  $\mathcal{N}_{\mathcal{NS}}$ . The weights are split in subfigures due to varying scales.

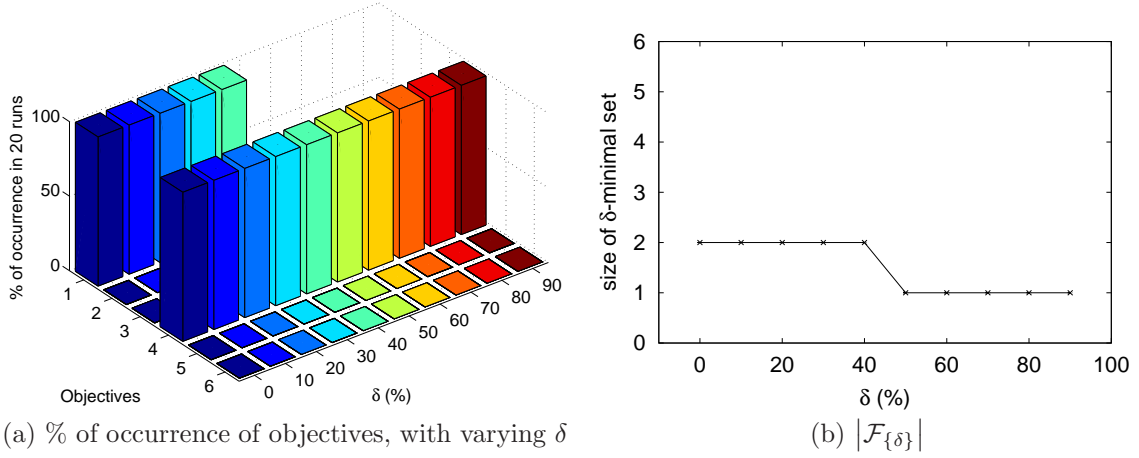


Figure 5.16: Work roll cooling design problem: Visual representation of the  $\delta$ -MOSS analysis, by the Approach-A (NL-MVU-PCA based framework), averaged over 20  $\mathcal{N}_{NS}$ .

### 5.7.3 Visual Representation of the Results

The information shared with the DM can be shown in an uncomplicated and time-efficient manner as represented in Figures 5.16 and 5.17, for  $\delta$ -MOSS and  $k$ -EMOSS analysis, respectively. While the former case gives a general idea of the problem redundancy, the latter informs the DM about the objective set of a specified size that yields minimal error.

### 5.7.4 Key Highlights of the Decision Support Offered by the Proposed Framework

The key highlights of the proposed decision support offered by the Approach-A, are as follows:

- The revelation of an *essential* objective set (Perspective-I) informs the DM that an objective set  $\mathcal{F}_s = \{f_1, f_4\}$  incurs an error of 0.612534% and therefore objectives  $\mathcal{F}_{redn} = \{f_2, f_3, f_5, f_6\}$  can be discarded. The identified essential objectives are related to the change in temperature at roll surface ( $f_1$ ), and radial stress at 9mm depth ( $f_4$ ).
- The preference-ranking of all the objectives (Perspective-II) reveals  $f_4 \triangleright f_1 \triangleright f_2 \triangleright f_5 \triangleright f_3 \triangleright f_6$ , where  $\triangleright$  denotes *more-important-than* relation, and assists with the following

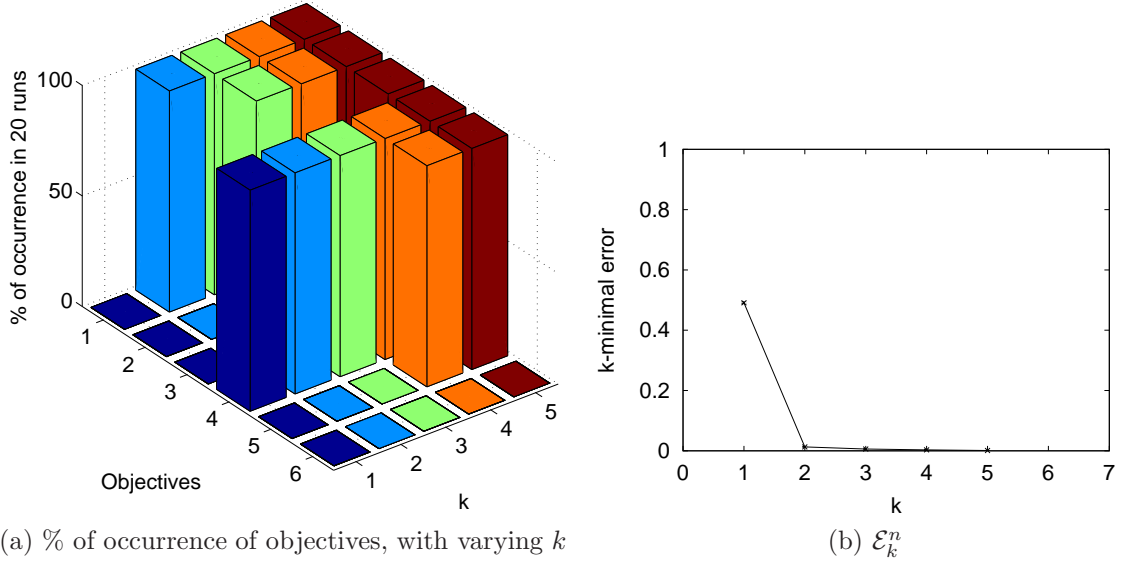


Figure 5.17: Work roll cooling design problem: Visual representation of the  $k$ -EMOSS analysis, by the Approach-A (NL-MVU-PCA based framework), averaged over 20  $\mathcal{N}_{\mathcal{S}}$ .

interpretation:

- The most important radial stress depth is 9mm, corresponding to objective  $f_4$ , while the most important change in temperature depth is at roll surface, corresponding to objective  $f_1$ .
  - Since  $f_4 \triangleright f_1$ , accounting for the radial stress is more critical than accounting for the change in temperature.
- If the DM allows for an error greater than 49.6501%, accounting for only a single objective, namely,  $f_4$  will suffice. On the other hand, if the DM wants to account for only two objectives which corresponds to the minimum error,  $f_1$  and  $f_4$  need to be accounted for.

## 5.8 Influence of the Used Distance Measure on the Accuracy of the Decision Support

The inconsistency in the decision support exhibited by Approach-A and that by Approach-B when applied to the radar waveform problem raises critical issues about the reliability of the decision support. Here it is argued that the reliability of a decision support would depend on the ability of the underlying objective reduction approaches to handle *nonlinearity* and *noise*. In this context *noise* relates to the difference in the dominance relation characterising the true POF and the solutions obtained from an MOEA, while *nonlinearity* relates to the nonlinear distribution (in objective space) of the non-dominated solutions obtained from an MOEA.

It is important to distinguish between linearity and nonlinearity of the non-dominated front and that of the objective functions. A non-dominated front will depend on how the given objectives covary, while the objective functions will depend how they vary across the search space. As a result:

- If all the objectives are linear and nonlinear but with the same degree of nonlinearity, then the resulting non-dominated front will be linear (Figure 5.18a).
- If the objectives have different degrees of nonlinearity, then the resulting non-dominated front will be nonlinear (Figure 5.18b).

For an objective reduction approach to evaluate a non-dominated solution set and apply  $\delta$ -MOSS and  $k$ -EMOSS analysis it is necessary to rely on a distance measure. Figure 5.18a shows that the use of  $L_1$  or  $L_2$  norm to measure distance between solutions does not account for nonlinearity, since the distribution of solutions between two solutions is not taken into consideration. For example, by considering the solution pairs  $(a, b)$  and  $(c, d)$ , if  $L_2$  norm is used, then the distance between the solutions is assumed to be equal, since  $L_2(a, b) = L_2(c, d) = D$ . However, the distribution of the intermediate solutions is not taken into

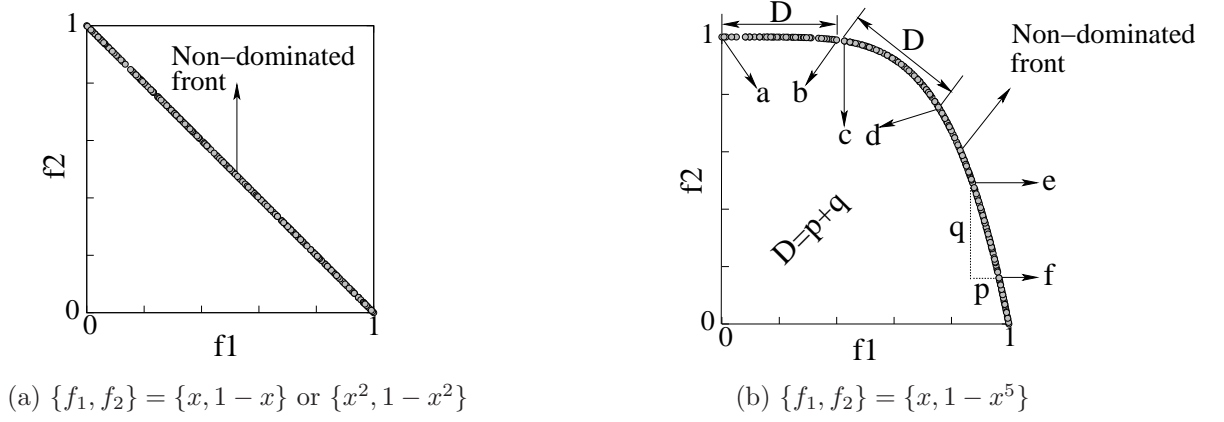


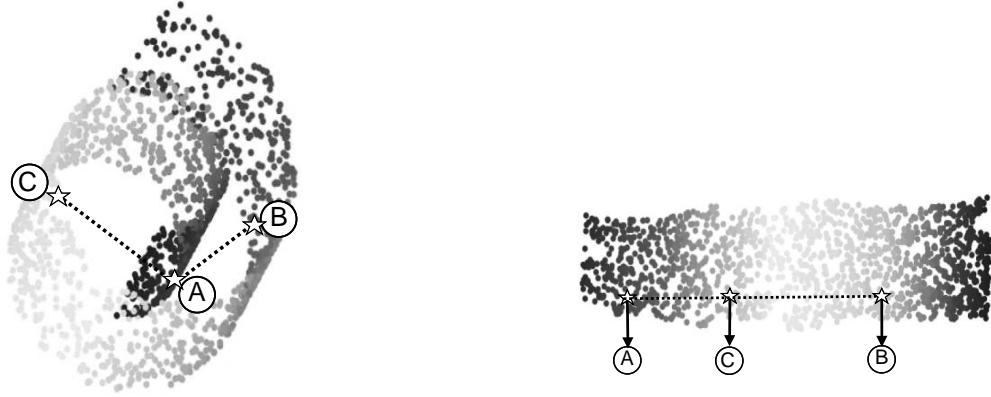
Figure 5.18: Illustration: (a) Two linear or identically nonlinear objectives result in a *linear* non-dominated front, and (b) two objectives with different degree of nonlinearity result in a *nonlinear* non-dominated front, where use of  $L_1$  or  $L_2$  norm will be inadequate to account for nonlinearity, and  $x \in [0, 1]$ .

account. The same situation holds for solution pairs  $(a, b)$  and  $(e, f)$  when  $L_1$  norm is used, since  $L_1(a, b) = L_1(e, f) = D$ . The previous discussion is used to highlight that an objective reduction approach that relies on  $L_1$  or  $L_2$  norm will be limited to linear objective reduction. An objective reduction approach with that limitation can only be used if the dataset is divided in clusters, such that in each cluster the non-dominated front is linear or if the dataset is unfolded.

### 5.8.1 Distance Measure in the Context of Approach-A

In the context of Approach-A the distance measure is referred by error measure and relates to the variance contribution of different objectives captured along the principal components (Equation 5.1). Consequently, the correctness of the error measured is dependent on the accuracy of the proposed objective reduction algorithms in determining the principal components of the non-dominated solution set. In Framework 2, the principal components are captured in the feature space, i.e. objective space, by L-PCA, or in the kernel space, i.e., unfolded feature space, by NL-MVU-PCA. The accuracy of the former case is limited by underlying assumptions that have been discussed in Section 2.3.2. While attempting to min-





(a)  $L_2(A, B) < L_2(A, C)$  erroneously infers  $B$  closer to  $A$ , than  $C$  is to  $A$

(b) Unfolding helps to restore the point-ordering, as over the manifold

Figure 5.19: Highlighting: (a) the limitations of  $L_2$  norm when the data occupies a low-dimensional submanifold, and (b) the remedy through unfolding (MVU).

imise the mean square error (Equation 2.9) if  $L_2$  norm is applied to Figure 5.19a, solution  $A$  is considered closer to  $B$  than to  $C$ . This is considered incorrect since, after unfolding the dataset (Figure 5.19b), solution  $A$  is indeed closer to  $C$  than to  $B$ .

The limitation above could be avoided if PCA is applied directly to the kernel space (Figure 5.19b). This highlights the superiority of NL-MVU-PCA over L-PCA when handling nonlinear, and as a result it is utilised to demonstrate the framework.

### 5.8.2 Distance Measure in the Context of Approach-B

The  $\delta$  error that is utilised by Approach-B is given by Equation 2.5, where: (i)  $A$  is a non-dominated set of solutions, (ii)  $\mathcal{F}$  is the problem original objective set, (iii)  $\mathcal{F}'$  is a subset of  $\mathcal{F}$  ( $\mathcal{F}' \subset \mathcal{F}$ ), and (iv)  $\delta_{max}$  is the maximum error that is incurred when  $x$  weakly dominates  $y$  ( $x, y \in A$ ) with respect to  $\mathcal{F}'$ . If  $\delta_{max} = 0$  it implies that the objectives in  $\mathcal{F} \setminus \mathcal{F}'$  are redundant. The limitations of Approach-B are inherit in the assumptions that Equation 2.5 is based on, as mentioned by the authors in [2]. Those assumptions are:

1. The  $\delta$  error across all the solutions ( $x, y \in A$ ) is comparable: It is assumed in [2] that *an error made close to the Pareto-optimal front is of the same importance as the same*

*error made far away from the Pareto-optimal front.* This can hold true *iff* the solutions on the non-dominated front are *equally* distributed. Otherwise, the importance of  $\delta$  in more dense regions which in general characterises more important regions, will be higher. As a result, when  $\delta$  is used to measure distances in situations where the density of solutions is not equal, it is likely that the error measured will be erroneous. This can happen if the distribution of solutions is like Gaussian (more midrange solutions than at extremes) or non-Gaussian (more extreme solutions than at midrange), which is then corroborated by the authors in [2] where they state (not necessarily in the same sequence) that *situations where the objective function values are not equally distributed are not considered in the study, for example: (i) situations where the decision maker prefers extremal solutions with maximal objective function values, or (ii) situations where the solutions close to extremal values are more unlikely than ones with midrange values.*

2.  $\delta$  error is comparable across all the objectives for any two given solutions ( $\max_{x,y \in A} \{\max_{f_i \in \mathcal{F}} \{f_i(x) - f_i(y)\}\}$ ): Another assumption made by the authors in [2] is that *all objective values have the same scale and reference point such that the small errors  $\delta$  are comparable among the objectives.* This assumption limits the application of the approach to linear objective reduction, since it can hold true only when all the objectives are either linear (Figure 5.20a) or identically nonlinear (same degree of nonlinearity Figure 5.20b). In Figure 5.20c, the objectives have different degree of nonlinearity and therefore the  $\delta$  error captured across the objectives is not comparable, along the non-dominated front. This is also corroborated by the authors in [2] by stating that *an incorporation of nonlinear objective functions would be extremely useful but remains future work.*

Due to the mentioned limitations, Approach-B is likely to fail when applied to instances such as in Figure 5.21a and 5.21b, where the non-dominated solutions are not equally distributed, and as a result the first assumption will be violated. It is also likely that Approach-B will fail when applied to the instance in Figure 5.21c, since the non-dominated front is not

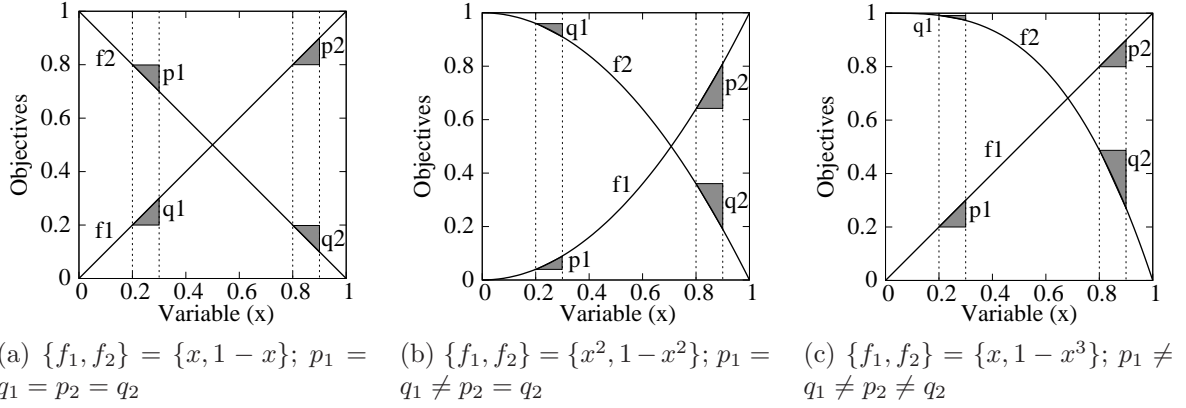


Figure 5.20: Illustration: Unless all the objectives are (a) linear, or (b) identically nonlinear, the DRP approach assumption of comparable  $\delta$  across all the objectives, over the entire non-dominated front will not hold. For all three cases,  $x \in [0, 1]$ .  $p_i$  and  $q_i$  show the variation in  $f_1$  and  $f_2$ , respectively, over different regions of the front.

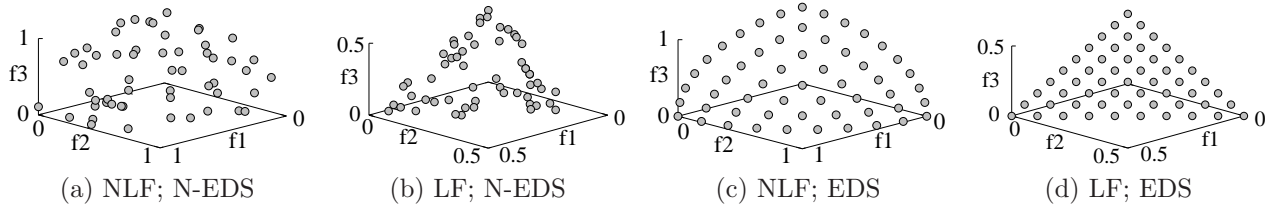


Figure 5.21: Highlighting the assumptions of the DRP approach, relating to the linearity of the front and equal distribution of solutions. LF and NLF stand for a linear and nonlinear front, respectively. EDS and N-EDS stand for equally distributed and not-equally distributed solutions, respectively.

linear, and as a result the second assumptions will be violated. In a case when the non-dominated front is linear and the solutions are equally distributed (as represented in Figure 5.21d) it is likely that Approach-B will work, given the assumptions that the approach is based on. This highlights the limitations and assumptions that Approach-B are based on, which perhaps explains the anonymity of its results for the radar waveform problem.

## 5.9 Comparison of the Efficacy of Approach-A and -B in Handling Noise

From the  $\delta$ -MOSS and  $k$ -EMOSS perspective this section investigates the efficacy of Approach-A vis-à-vis Approach-B in handling solution sets with a *noisy* POF-approximation, meaning that, their dominance relations are different from those of the true POF. Solution sets that are likely to provide a *noisy* POF-approximation are those generated by an MOEA due to their stochastic nature. Since the dominance relations of noisy solution sets are different from those on the true POF, to reveal the preference-structure of the objectives on the true POF using an objective reduction approach based on preserving the dominance relations is itself questionable. The terms POF-approximation and POF-representation are differentiated in Section 4.8.4, where a solution set with a correlation-structure equal to the one obtained on the POF is inferred as having a good POF-representation, and:

- I1 A realistic solution set (obtained by an MOEA) characterised by a poor POF-approximation might reveal a good POF-representation, since a good POF-representation is not necessary a good POF-approximation, as stated in Section 4.8.4 from Chapter 4.
- I2 An objective reduction approach based on preserving the correlation-structure provides more accurate results than another approach based on preserving the dominance relations, when applied to realistic solution sets.

To investigate the above arguments in the context of  $\delta$ -MOSS and  $k$ -EMOSS analysis, the test problems (namely DTLZ5( $I, M$ ) and DTLZ2<sub>BZ</sub>) characteristics are as follows:

- For DTLZ5( $I, M$ ), where  $M$  represents the number of objectives, and  $I$  ( $I \leq M$ ) denotes the dimensionality of the POF: (i) the objectives in  $\{f_1, \dots, f_{M-I+1}\}$  are non-conflicting among themselves and (ii) each objective in  $\{f_{M-I+2}, \dots, f_M\}$  is in conflict with all the remaining objectives, hence:

- The  $\delta$ -MOSS analysis should identify:  $\mathcal{F}_{\{\delta=0\}s} = \{f_{M-I+1}, \dots, f_M\}$ .
- The  $k$ -EMOSS analysis should identify:  $\mathcal{E}_{k=I}^n = 0$  corresponding to  $\mathcal{F}_{\{k=I\}s} = \{f_{M-I+1}, \dots, f_M\}$ .
- For DTLZ2<sub>BZ</sub>( $M$ ), where  $M$  represents the number of objectives, each objective is in conflict with the rest, hence:
  - The  $\delta$ -MOSS analysis should identify:  $\mathcal{F}_{\{\delta=0\}s} = \{f_1, \dots, f_M\}$  and with a linear increase in  $\delta$  from 0 to 1, the  $|\mathcal{F}_{\{\delta\}s}|$  should linearly decrease.
  - The  $k$ -EMOSS analysis should identify:  $\mathcal{E}_{k=M}^n = 0$  corresponding to  $\mathcal{F}_{\{k=M\}s} = \{f_1, \dots, f_M\}$  and with a linear increase in  $k$ , the  $\mathcal{E}_k^n$  should linearly decrease.

The argument in Item I1 could be realised in Figure 5.22, where a comparison is made between  $\mathcal{N}_{\mathcal{P}}$  and  $\mathcal{N}_{\epsilon}$ . It can be noticed that  $\mathcal{N}_{\epsilon}$  fails to achieve a complete convergence towards the POF (Figure 5.22a vis-à-vis Figure 5.22b) and diversity across the POF (Figure 5.22c vis-à-vis Figure 5.22d). In that:

- $\mathcal{N}_{\epsilon}$  for DTLZ5(2,10): fails to accurately capture the relationships between all the objectives. For example, while each objective pair in  $f_1$ – $f_9$  are non-conflicting in  $\mathcal{N}_{\mathcal{P}}$ , for  $\mathcal{N}_{\epsilon}$  some solutions show conflict between them (e.g., the crossing lines between  $f_6$  and  $f_7$  in Figure 5.22b). However, as the fraction of such misleading solutions is reasonable small, the global correlation–structure remains the same, implying a good POF–representation by  $\mathcal{N}_{\epsilon}$ .
- $\mathcal{N}_{\epsilon}$  for DTLZ2<sub>BZ</sub>(10): captures the relationships between all the objectives, as in  $\mathcal{N}_{\mathcal{P}}$  (all objectives are in conflict). Hence, the global correlation–structure will be the same as in  $\mathcal{N}_{\mathcal{P}}$ , implying a good POF–representation by  $\mathcal{N}_{\epsilon}$ .

The above two sample problem observations validates the argument in Item I1, but now the focus is in Item I2 argument. Towards it, the  $\delta$ -MOSS and  $k$ -EMOSS analysis is applied

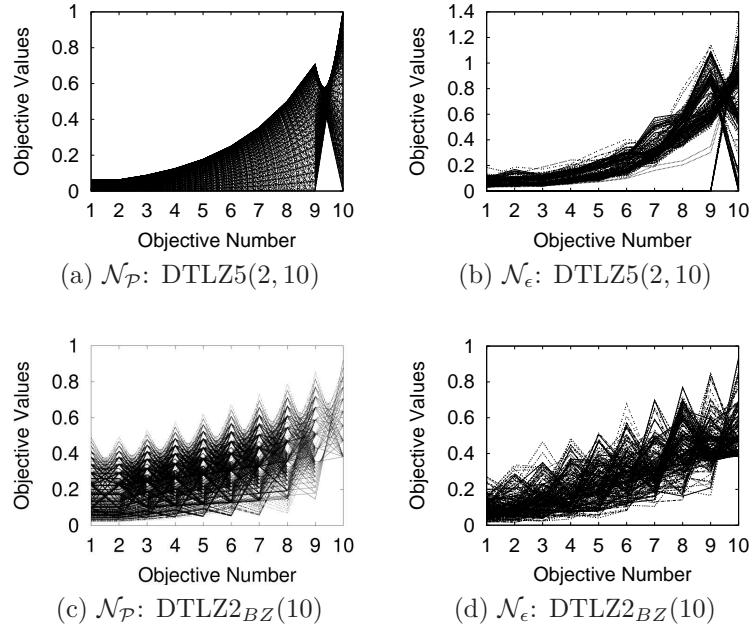


Figure 5.22: Comparison between solution generated on the true POF (referred to as  $\mathcal{N}_{\mathcal{P}}$ ) and solutions obtained by  $\epsilon$ -MOEA (referred to as  $\mathcal{N}_{\epsilon}$ ) (one run): Parallel coordinate plots for DTLZ5(2, 10) and DTLZ2<sub>BZ</sub>(10), highlighting the *noise* in  $\mathcal{N}_{\epsilon}$  as opposed to  $\mathcal{N}_{\mathcal{P}}$ .

to the same test problems and it is captured in Figures 5.23 and 5.24, for DTLZ5(2, 10) and DTLZ2<sub>BZ</sub>(10), respectively:

- Based on DTLZ5(2, 10) the following can be noted in Figure 5.23:
  - For  $\mathcal{N}_{\mathcal{P}}$ : both Approach-A and Approach-B accurately identify  $|\mathcal{F}_{\{\delta=0\}s}| = 2$  (Figure 5.23a) and  $\mathcal{E}_{k=I}^n = 0$  (Figure 5.23b).
  - For  $\mathcal{N}_{\epsilon}$ : Approach-A is able to reproduce the results as with  $\mathcal{N}_{\mathcal{P}}$ , but Approach-B fails, in that, it wrongly identifies  $|\mathcal{F}_{\{\delta=0\}s}| = 10$  (Figure 5.23c), and  $\mathcal{E}_{k=I}^n \approx 17\%$  (Figure 5.23d).
- Based on DTLZ2<sub>BZ</sub>(10) the following can be noted in Figure 5.24:
  - For  $\mathcal{N}_{\mathcal{P}}$ : both Approach-A and Approach-B accurately identify  $\mathcal{F}_{\{\delta=0\}s} = 10$  (Figure 5.24a) and  $\mathcal{E}_{k=10}^n = 0$  (Figure 5.24b).

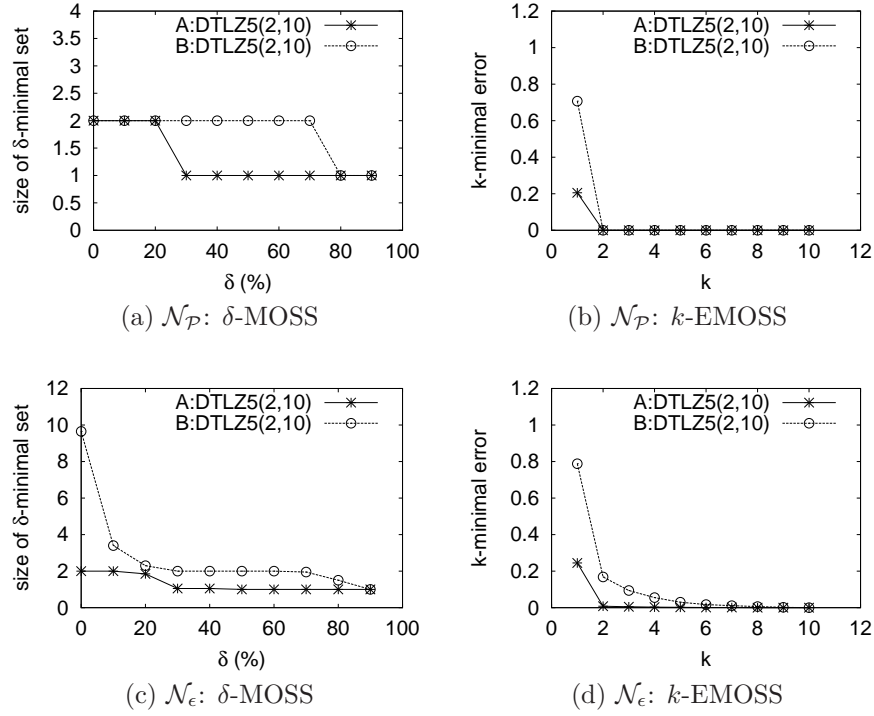


Figure 5.23: A comparative view on Approach-A vis-à-vis Approach-B for  $\delta$ -MOSS and  $k$ -EMOSS analysis applied to DTLZ5(2, 10), based on  $\mathcal{N}_{\mathcal{P}}$  and  $\mathcal{N}_{\epsilon}$ .

- For  $\mathcal{N}_{\epsilon}$ : Approach-A is able to reproduce the results as with  $\mathcal{N}_{\mathcal{P}}$ , but Approach-B fails as it wrongly identifies  $\mathcal{F}_{\{\delta=0\}s} = 8$ . Furthermore, in the case of Approach-A, while a gradual increase in  $\delta$  leads to a gradual increase in  $|\mathcal{F}_{\{\delta\}s}|$ , and a gradual decrease in  $k$  leads to a gradual increase in  $|\mathcal{E}_k^n|$ ; the variation in case of Approach-B is quite abrupt.

The argument in Item I2 is validated by the above observations. Furthermore, for different versions of DTLZ5( $I, M$ ) and DTLZ2<sub>BZ</sub>( $M$ ) problems, the results are summarised in Tables 5.14 and 5.15, for  $\mathcal{N}_{\mathcal{P}}$  and  $\mathcal{N}_{\epsilon}$ , respectively. The results show the same trend and confirms higher accuracy by Approach-A over Approach-B. Hence, a decision support based on preserving the correlation-structure is likely to be more reliable than one based on preserving the dominance relations. Finally, the reasons behind the anonymity of the results obtained for the radar waveform problem in Section 5.4 for Approach-B, could perhaps be explained by dominance relations in  $\mathcal{N}_{\mathcal{MS}}$ —different from those on the underlying POF.

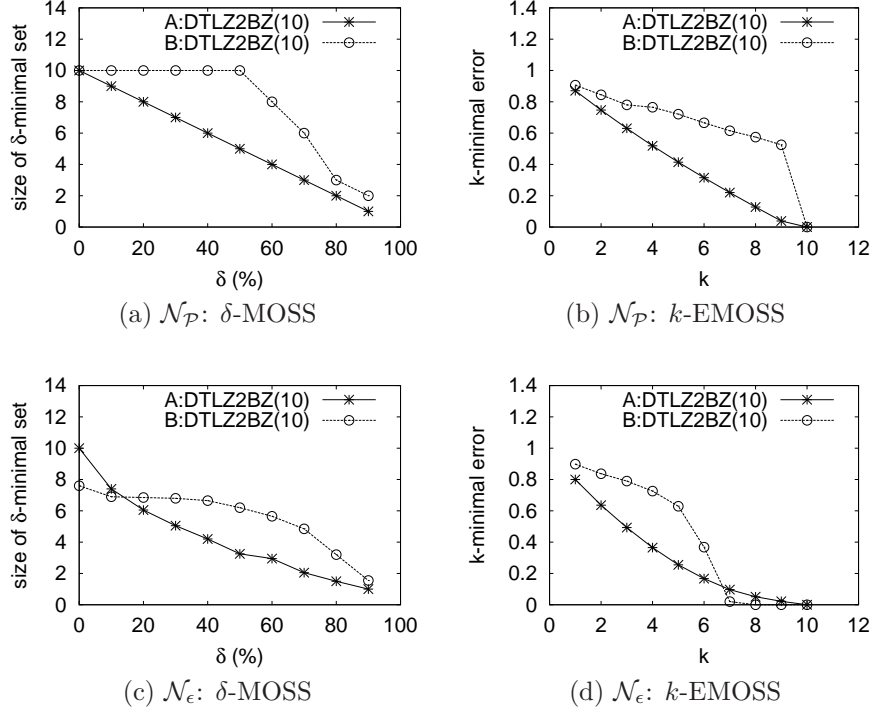


Figure 5.24: A comparative view on Approach-A vis-à-vis Approach-B for  $\delta$ -MOSS and  $k$ -EMOSS analysis applied to  $\text{DTLZ2}_{BZ}(10)$ , based on  $\mathcal{N}_{\mathcal{P}}$  and  $\mathcal{N}_{\epsilon}$ .

Table 5.14:  $\delta$ -MOSS and  $k$ -EMOSS analysis for the redundant  $\text{DTLZ5}(I, M)$  and non-redundant  $\text{DTLZ2}_{BZ}$  problems, based on  $\mathcal{N}_{\mathcal{P}}$

This table compares Approach-A (results in columns) vis-à-vis Approach-B (results in brackets adjacent to each column element)

Problems	$ \mathcal{F}_{\{\delta\}} $ : $\delta$ -MOSS ( $\delta \in [0 : 1]$ )			$\mathcal{E}_k^n$ : $k$ -EMOSS ( $k = \lceil pM \rceil, p \in [0 : 1]$ )		
	$\delta = 0.0$	$\delta = 0.2$	$\delta = 0.4$	$k = I$	$p = 0.6$	$p = 0.9$
DTLZ5(2, 05)	2 (2)	1 (2)	1 (2)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
DTLZ5(2, 10)	2 (2)	2 (2)	1 (2)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
DTLZ5(2, 50)	2 (2)	2 (2)	1 (2)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
DTLZ5(3, 05)	3 (3)	2 (3)	2 (3)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
DTLZ5(3, 20)	3 (3)	2 (3)	2 (3)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
DTLZ5(5, 10)	5 (5)	4 (5)	3 (5)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
DTLZ5(5, 20)	5 (5)	4 (5)	3 (5)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
DTLZ5(7, 10)	7 (7)	5 (7)	4 (7)	0.000 (0.000)	0.059 (0.707)	0.000 (0.000)
DTLZ5(7, 20)	7 (7)	5 (7)	4 (7)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Problems	$\delta = 0.0$	$\delta = 0.2$	$\delta = 0.4$	$p = 0.3$	$p = 0.6$	$p = 0.9$
DTLZ2 <sub>BZ</sub> (05)	05 (05)	04 (05)	03 (05)	0.574 (0.635)	0.365 (0.576)	0.000 (0.000)
DTLZ2 <sub>BZ</sub> (10)	10 (10)	08 (10)	06 (10)	0.630 (0.551)	0.315 (0.470)	0.038 (0.371)
DTLZ2 <sub>BZ</sub> (15)	15 (15)	11 (15)	07 (14)	0.528 (0.509)	0.261 (0.371)	0.015 (0.270)



Table 5.15:  $\delta$ -MOSS and  $k$ -EMOSS analysis for the redundant DTLZ5( $I, M$ ) and non-redundant DTLZ2<sub>BZ</sub> problems, averaged over 20 different samples of  $\mathcal{N}_\epsilon$ 

The results represent the *Mean  $\pm$  Standard deviation* for  $|\mathcal{F}_{\{\delta\}}|$  and  $\mathcal{E}_k^n$ , in their respective cases. The columns elements correspond to Approach-A, while those in adjacent brackets correspond to Approach-B  
 (a)  $|\mathcal{F}_{\{\delta\}}|$ :  $\delta$ -MOSS ( $\delta \in [0 : 1]$ )

Problems	$\delta = 0.0$	$\delta = 0.2$	$\delta = 0.4$
DTLZ5(2, 05)	2.00 $\pm$ 0.00 (05.00 $\pm$ 0.00)	1.90 $\pm$ 0.31 (2.60 $\pm$ 0.60)	1.00 $\pm$ 0.00 (2.00 $\pm$ 0.00)
DTLZ5(2, 10)	2.00 $\pm$ 0.00 (09.65 $\pm$ 0.75)	1.85 $\pm$ 0.37 (2.30 $\pm$ 0.47)	1.05 $\pm$ 0.22 (2.00 $\pm$ 0.00)
DTLZ5(2, 50)	2.00 $\pm$ 0.00 (10.90 $\pm$ 1.21)	2.00 $\pm$ 0.00 (7.05 $\pm$ 1.00)	1.05 $\pm$ 0.22 (3.90 $\pm$ 0.85)
DTLZ5(3, 05)	3.00 $\pm$ 0.00 (05.00 $\pm$ 0.00)	2.00 $\pm$ 0.00 (3.05 $\pm$ 0.22)	2.00 $\pm$ 0.00 (3.00 $\pm$ 0.00)
DTLZ5(3, 20)	3.00 $\pm$ 0.00 (10.60 $\pm$ 1.27)	2.15 $\pm$ 0.37 (5.30 $\pm$ 0.80)	2.00 $\pm$ 0.00 (4.00 $\pm$ 0.32)
DTLZ5(5, 10)	5.20 $\pm$ 0.41 (08.85 $\pm$ 0.81)	4.00 $\pm$ 0.00 (5.00 $\pm$ 0.00)	3.00 $\pm$ 0.00 (5.00 $\pm$ 0.00)
DTLZ5(5, 20)	5.45 $\pm$ 0.51 (10.35 $\pm$ 1.09)	4.00 $\pm$ 0.00 (6.00 $\pm$ 0.56)	3.00 $\pm$ 0.00 (5.40 $\pm$ 0.50)
DTLZ5(7, 10)	7.45 $\pm$ 0.60 (08.40 $\pm$ 0.94)	5.25 $\pm$ 0.44 (7.00 $\pm$ 0.00)	4.00 $\pm$ 0.00 (7.00 $\pm$ 0.00)
DTLZ5(7, 20)	7.85 $\pm$ 0.75 (10.70 $\pm$ 0.92)	5.25 $\pm$ 0.44 (7.60 $\pm$ 0.50)	4.00 $\pm$ 0.00 (7.30 $\pm$ 0.47)
Problems	$\delta = 0.0$	$\delta = 0.2$	$\delta = 0.4$
DTLZ2 <sub>BZ</sub> (05)	05.00 $\pm$ 0.00 (05.00 $\pm$ 0.00)	04.00 $\pm$ 0.00 (05.00 $\pm$ 0.00)	03.00 $\pm$ 0.00 (05.00 $\pm$ 0.00)
DTLZ2 <sub>BZ</sub> (10)	10.00 $\pm$ 0.00 (07.60 $\pm$ 0.75)	06.05 $\pm$ 0.22 (06.85 $\pm$ 0.49)	04.20 $\pm$ 0.41 (06.65 $\pm$ 0.49)
DTLZ2 <sub>BZ</sub> (15)	15.00 $\pm$ 0.00 (08.55 $\pm$ 1.00)	06.65 $\pm$ 0.49 (07.25 $\pm$ 0.55)	04.60 $\pm$ 0.50 (06.95 $\pm$ 0.39)

(b)  $\mathcal{E}_k^n$ :  $k$ -EMOSS ( $k = \lceil pM \rceil, p \in [0 : 1]$ )

Problems	$k = I$	$p = 0.6$	$p = 0.9$
DTLZ5(2, 05)	0.008 $\pm$ 0.005 (0.213 $\pm$ 0.050)	0.004 $\pm$ 0.002 (0.144 $\pm$ 0.042)	0.000 $\pm$ 0.000 (0.000 $\pm$ 0.000)
DTLZ5(2, 10)	0.008 $\pm$ 0.004 (0.168 $\pm$ 0.065)	0.001 $\pm$ 0.001 (0.017 $\pm$ 0.007)	0.000 $\pm$ 0.000 (0.002 $\pm$ 0.002)
DTLZ5(2, 50)	0.020 $\pm$ 0.015 (0.194 $\pm$ 0.127)	0.001 $\pm$ 0.001 (0.000 $\pm$ 0.000)	0.000 $\pm$ 0.000 (0.000 $\pm$ 0.000)
DTLZ5(3, 05)	0.001 $\pm$ 0.001 (0.125 $\pm$ 0.049)	0.001 $\pm$ 0.001 (0.125 $\pm$ 0.049)	0.000 $\pm$ 0.000 (0.000 $\pm$ 0.000)
DTLZ5(3, 10)	0.007 $\pm$ 0.006 (0.164 $\pm$ 0.044)	0.001 $\pm$ 0.000 (0.001 $\pm$ 0.001)	0.000 $\pm$ 0.000 (0.000 $\pm$ 0.000)
DTLZ5(5, 10)	0.003 $\pm$ 0.006 (0.078 $\pm$ 0.045)	0.001 $\pm$ 0.000 (0.016 $\pm$ 0.013)	0.000 $\pm$ 0.000 (0.000 $\pm$ 0.001)
DTLZ5(5, 20)	0.013 $\pm$ 0.018 (0.086 $\pm$ 0.055)	0.000 $\pm$ 0.000 (0.000 $\pm$ 0.000)	0.000 $\pm$ 0.000 (0.000 $\pm$ 0.000)
DTLZ5(7, 10)	0.011 $\pm$ 0.013 (0.025 $\pm$ 0.040)	0.078 $\pm$ 0.017 (0.608 $\pm$ 0.090)	0.000 $\pm$ 0.000 (0.001 $\pm$ 0.002)
DTLZ5(7, 20)	0.015 $\pm$ 0.013 (0.025 $\pm$ 0.048)	0.000 $\pm$ 0.000 (0.000 $\pm$ 0.000)	0.000 $\pm$ 0.000 (0.000 $\pm$ 0.000)
Problems	$p = 0.3$	$p = 0.6$	$p = 0.9$
DTLZ2 <sub>BZ</sub> (05)	0.456 $\pm$ 0.020 (0.859 $\pm$ 0.039)	0.266 $\pm$ 0.017 (0.785 $\pm$ 0.018)	0.000 $\pm$ 0.000 (0.000 $\pm$ 0.000)
DTLZ2 <sub>BZ</sub> (10)	0.493 $\pm$ 0.033 (0.790 $\pm$ 0.045)	0.167 $\pm$ 0.019 (0.368 $\pm$ 0.185)	0.022 $\pm$ 0.005 (0.000 $\pm$ 0.000)
DTLZ2 <sub>BZ</sub> (15)	0.298 $\pm$ 0.021 (0.718 $\pm$ 0.087)	0.044 $\pm$ 0.007 (0.065 $\pm$ 0.188)	0.003 $\pm$ 0.000 (0.000 $\pm$ 0.000)

## 5.10 Scope of the Decision Support Offered by the Proposed Framework

The benefits of the decision support have been demonstrated for those MaOPs, where not all the objectives are *essential* to describe the POF and/or not all the *essential* objectives are equally important. For problems where all objectives are *essential* and equally important, the captured objective preference-weights by the decision support will be almost identical, in which case it will not be useful for the DM. For handling such problems:

- one may need to rely on the DM preferences that are embroiled in *subjectivity* (despite their disadvantages), if the MCDM based MOEAs are to be employed;
- efforts can be made in improving the computational efficiency of the otherwise better performing (in terms of POF–approximation) indicator based MOEAs or a decomposition based MOEA/D;
- the inefficiency of existing concepts of Pareto-optimality and Pareto-dominance in modelling and simulating human decision-making needs to be countered. As highlighted in [71], the notion of *optimality* is captured in a narrow sense, since when two solutions are compared: (i) a solution that is marginally better in one objective out of many qualifies as non-dominated (weakens the pressure towards convergence to the POF) and (ii) some crucial aspects in human decision-making are not taken into consideration — the number of improved or equal objective values, and the size of such improvements.

Despite the above, it is common for real-world problems to be characterised by some redundant objectives and different importance between the objectives. As demonstrated for four real-world problems, the decision support proved to be immensely useful by articulating the DM’s preferences with *objectivity*, *repeatability*, *consistency* and *coherence*.

## 5.11 Summary

The difficulties faced by existing MOEAs in dealing with MaOPs have stimulated a growing interest in MCDM based MOEAs. These algorithms rely on the DM's preferences to guide the search process towards a few solutions, rather than the whole POF. However, as suggested by research in psychology and cognitive sciences, people are neither very good at handling large amounts of data or expressing them in quantified ways, nor at handling more than several ( $7 \pm 2$ ) factors at the time. The identified limitations could lead to suboptimal or biased solutions when dealing with MaOPs. Given this, it is identified that the DM's preferences when applied to MaOPs may be characterised by lack of: (i) *objectivity* (a rational basis); *repeatability* (identical preferences for identical options); *consistency* (alike preferences across multiple interaction stages); and *coherence* (alike preferences by multiple DMs).

To counter the above limitations, a machine learning based framework has been proposed in this chapter that aims to guide the DM's preferences with rationality. To achieve this purpose, the proposed decision support is based on the premise that a logical implementation can be achieved through learning the preference–structure of the objective functions, inherent in the formulation of the optimisation problem. The rationale behind this framework is based on the fact the modelling is generally pre-given by the joint expertise of engineers and technical managers whose knowledge is embedded into the problem structure. That same problem structure can be revealed to the DM and help him/her to articulate the objective preferences better. Towards this, the proposed framework is an extension of Framework 1 (from Chapter 4) which is able to:

1. reveal the smallest set of conflicting (essential) objectives which can generate the same solution set as the original problem;
2. preference-rank all the objectives;

3. reveal the smallest objective sets corresponding to pre-specified  $\delta$  ( $0 \leq \delta \leq 1$ ) errors ( $\delta$ -MOSS);
4. reveal the objective sets of pre-specified size  $k$  that correspond to minimum error ( $k$ -EMOSS); and
5. present the information to the DM in an uncomplicated, time-efficient, visualisation scheme.

The utility of the proposed framework has been evaluated using a wide variety of test problems and four real-world problems. A comparative analysis between the proposed framework and an alternative approach based on dominance relation presentation (DRP) in handling *nonlinearity* and *noise* has been conducted. In this context *noise* relates to the difference in the dominance relations characterising the true POF and the solutions obtained from an MOEA, while *nonlinearity* relates to the nonlinear distribution (in objective space) of the non-dominated solutions obtained from an MOEA. To this end, NL-MVU-PCA is referred to as Approach-A and the exact algorithm from DRP is referred to as Approach-B. For Approach-A, NL-MVU-PCA is chosen over L-PCA due to its superiority in terms of handling data sets with different distributions (Section 4.14.1 in Chapter 4). For Approach-B, the exact is chosen over the greedy algorithm since the latter does not guarantee a 0-minimum objective set.

For a particular real-world problem, namely radar waveform (Section 5.4), Approach-B has revealed contrasting trends to those predicted by the author in [162], while the obtained results by Approach-A can be considered more reliable in regards to the predictions. Since both decision supports have revealed inconsistency in the provided decision support, it is argued that the reliability of the decision supports would depend on their ability to handle *nonlinearity* and *noise*. To this extent, for an objective reduction approach to conduct  $\delta$ -MOSS and  $k$ -EMOSS it needs to rely on a distance measure as follows:

- In the context of Approach-A the distance measure is referred to by error measure and

relates to the variance contribution of different objectives captured along the principal components.

- In the context of Approach-B the distance measure is based on the definition of  $\delta_{max}$  and the assumptions that are made limit the application of the approach to: (i) equally distributed solutions in the non-dominated front and (ii) to linear or identically non-linear objectives. Hence, it is likely that Approach-B limitations could perhaps explain the anonymity of the results on the radar waveform problem.

Moreover, in Chapter 4 the PCA-based and DRP-based objective reduction algorithms have been compared in finding a  $\delta$ -minimum objective set for  $\delta = 0$ . In this chapter the comparative analysis between the two approaches is extended for: (i)  $\delta$ -MOSS analysis where  $0 \leq \delta \leq 1$  and (ii)  $k$ -EMOSS analysis where  $k = \lceil pM \rceil$  and  $p$  ( $0 \leq p \leq 1$ ). Based on the results obtained the following remarks can be made:

- Approach-A and Approach-B have both accurately identified the size of the  $\delta$ -minimal set and the  $k$ -minimal set when applied to solution sets without noise ( $\mathcal{N}_{\mathcal{P}}$ ). This highlights that Approach-B is likely to provide accurate results if the solution set is not characterised by noise.
- For solution sets characterised by noise ( $\mathcal{N}_{\epsilon}$ ), Approach-A was able to reproduce the results as with  $\mathcal{N}_{\mathcal{P}}$  but Approach-B has failed. This highlights the superiority of Approach-A over Approach-B in handling solution sets with *noise* in the context of  $\delta$ -MOSS and  $k$ -EMOSS analysis.

In resume, the obtained results have revealed the superiority of the Approach-A over Approach-B in handling *nonlinearity* and *noise*. Hence, a decision support based on preserving the correlation-structure is likely to be more reliable than one preserving the dominance relations if the underlying solution set is characterised by *noise*.

## Chapter 6

# *Online* Objective Reduction Framework

In Chapter 4, the proposed objective reduction algorithms (L-PCA and NL-MVU-PCA) have been implemented in a framework (Framework 1) that is *offline* and iterative in nature. The *offline* feature relates to the fact that these algorithms are applied after an MOEA run, while the iterative feature implies that only the objectives identified as *essential* in a particular iteration are considered in the next iteration. Besides their advantages, these features may pose some limitations in the wake of the following:

- in situations, where the intervention of the DM is either not possible or not advisable, an *online* implementation may be desired.
- when  $M \gg 4$ , the POF-approximation/representation in the first iteration may be too poor, and an *essential* objective may erroneously be eliminated as redundant, with no scope of being reconsidered in subsequent iterations.
- in problems where the relationships (conflict or non-conflict) between the objectives may vary across the search space, the objectives inferred as redundant in the initial iterations may actually be *essential* with regard to the true POF, and vice-versa.

Notably, the critical limitation of the Framework 1 emanates from the manner it iterates, in that, an objective eliminated in a particular iteration has no scope of being re-inducted

in subsequent iterations, potentially leading to inaccurate deductions. In this background, this chapter implements the objective reduction algorithms in an *online* framework, with an improvisation that at any instant, all the objectives are retained with different probabilities. Towards the proof-of-the-concept, several many-objective test problems are used as a reference to demonstrate that the probabilistic retention of all the objectives serves as a self-correcting mechanism, leading to enhanced accuracy of the *online* framework over its *offline* counterpart.

## 6.1 *Online* Objective Reduction Framework

This section presents an *online* objective reduction framework that aims to counter the limitations of the *offline* Framework 1 associated with the potent scenario of an *essential* objective being erroneously eliminated, and having no scope of being re-considered in subsequent iterations. Before introducing the *online* framework here, it is important to put the frameworks proposed in the previous chapters, into perspective. Towards it, a distinction needs to be made between the objective reduction algorithms (L-PCA and NL-MVU-PCA) and the frameworks developed around them through post-processing of the knowledge obtained from these algorithms. This calls for re-visiting the following:

1. the *offline* Framework 1 for objective reduction, in Chapter 4: here, the objective reduction algorithms were applied to an MOEA population corresponding to a pre-specified objective set, to determine an *essential* objective set ( $\mathcal{F}_s$ ) and the errors associated with the omission of each of the redundant objectives ( $\mathcal{E}_i \equiv f_i \in \mathcal{F}_{redn}$ ). If the pre-specified objective set and  $\mathcal{F}_s$  happened to be the same, the procedure was terminated, else repeated with  $\mathcal{F}_s$  as the next pre-specified objective set. In this case, one iteration of the framework was nothing but an application of the objective reduction algorithms once, followed by the decision to either terminate or repeat the procedure.
2. the Framework 2 for decision support, in Chapter 5: here, the objective reduction

algorithms were applied to an MOEA population corresponding to a pre-specified objective set, to determine an  $\mathcal{F}_s$  and the errors associated with the omission of *all*<sup>1</sup> the objectives, i.e., each of the *essential* and redundant ones ( $\mathcal{E}_i \equiv f_i, i = 1, \dots, M$ ). These errors when normalised led to the preference-weights ( $w_i = \mathcal{E}_i^n, i = 1, \dots, M$ ), whose post-processing led to the decision support in terms of the  $\delta$ -MOSS and  $k$ -EMOSS analysis. In this case, one iteration of the framework was nothing but an application of the objective reduction algorithms once, followed by the post-processing of results leading to the  $\delta$ -MOSS and  $k$ -EMOSS analysis.

The *online* framework to be proposed in this chapter, is being based on a similar structure as in Frameworks 1 and 2, and that includes:

- objective reduction module that involves an application of the objective reduction algorithms;
- post-processing module that involves application of the proposed probabilistic objective selection scheme.

### 6.1.1 Objective Reduction

As discussed above, the objective reduction algorithms (L-PCA/NL-MVU-PCA) were so merged in the earlier proposed frameworks<sup>2</sup>, that neither their scope nor their implementation steps were distinctively established vis-à-vis that of the respective frameworks. For instance, in Framework 1, L-PCA/NL-MVU-PCA were used to determine the errors associated with the omission of the redundant objectives (besides  $\mathcal{F}_s$ ), while in Framework 2, they were used to determine the errors associated with the omission of *all* the objectives (besides  $\mathcal{F}_s$ ) which post-normalisation were treated as preference-weights for the objectives. For the

---

<sup>1</sup>This is unlike in Framework 1, where only the errors associated with the omission of redundant objectives were determined.

<sup>2</sup>These include the *offline* Framework 1 and the Framework 2 for decision support.



sake of clarity, this chapter summarises the scope and implementation steps of the objective reduction algorithms in Algorithm 3.

---

**Algorithm 3:** Objective Reduction Algorithms
 

---

**Input:**  
 $\mathcal{F}_0$ : Initial objective set.  
 $\mathcal{H}$ : set of objective vectors of the non-dominated solutions corresponding to  $\mathcal{F}_0$ .

```

1 begin
2   Based on  $\mathcal{H}$  compute a positive semi-definite matrix:
3   – Correlation matrix  $R$  (Equation 2.11) for linear objective reduction algorithm (L-PCA).
4   – Kernel matrix  $K$  (Equation 2.15) for nonlinear objective reduction algorithm (NL-MVU-PCA).
5   Compute the eigenvalues and eigenvectors (principal components) of  $R$  or  $K$  (Section 4.1.2).
6   Perform the eigenvalue analysis (Section 4.1.3) to identify the set of important objectives  $\mathcal{F}_e \subseteq \mathcal{F}_0$ .
7   Perform the RCM analysis (Section 4.1.4) to identify the identically correlated subsets ( $\mathcal{S}$ ) in  $\mathcal{F}_e$ .
8   if identically correlated subsets ( $\mathcal{S}$ ) do not exist then
9     | Stop and declare  $\mathcal{F}_e$  as  $\mathcal{F}_s$ —the essential objective set.
10  end
11  else
12    | Apply the selection scheme (Section 4.1.5) to identify the most significant objective in each  $\mathcal{S}$ , to arrive at  $\mathcal{F}_s$ , such that  $\mathcal{F}_s \subseteq \mathcal{F}_e$ .
13  end
14  Compute the error  $\mathcal{E}_i$  associated with each objective using Equation 5.1.
15  Declare the redundant objective set as  $\mathcal{F}_{redn} = \mathcal{F}_0 \setminus \mathcal{F}_s$ , and error associated with the omission of  $\mathcal{F}_{redn}$  be given by  $\mathcal{E}_{redn} = \sum_{i: f_i \in \mathcal{F}_{redn}} \mathcal{E}_i$ .
16  Declare the preference-weights ( $w_i$ ) for all the objectives (Equation 5.4).
17 end
```

---

Notably, the *online* framework to be proposed will be based on post processing of the *essential* objective set ( $\mathcal{F}_s$ ) and the preference-weights ( $w_i$ s) determined by Algorithm 3.

### 6.1.2 Post-processing: Probabilistic Objective Selection Scheme

This section presents a scheme that post processes the information ( $\mathcal{F}_s$  and  $w_i$ s) revealed by Algorithm 3, to probabilistically determine objective sets that an MOEA would handle.

While the discussion on how such a probabilistic objective set will be utilised in the *online* framework is deferred until the next section, the details of the scheme are as follows:

1. let the collective preference-weights for  $\mathcal{F}_s$  and  $\mathcal{F}_{redn}$  be given by  $\mathcal{W}_s$  and  $\mathcal{W}_{redn}$ , respectively, as defined below:

$$\left. \begin{aligned} \mathcal{W}_s &= \sum_{j: f_j \in \mathcal{F}_s} w_j, \\ \mathcal{W}_{redn} &= \sum_{j: f_j \in \mathcal{F}_{redn}} w_j \end{aligned} \right\} \quad (6.1)$$

2. let the preference-weights for the objectives (both in  $\mathcal{F}_s$  and  $\mathcal{F}_{redn}$ ), sorted by magnitude in descending order be given by  $w_{si}$  ( $w_{s1} \geq \dots \geq w_{sM}$ ) and the corresponding objectives by  $\{f_{s1}, \dots, f_{sM}\}$ .
3. let the confidence level in the accuracy of the  $\mathcal{F}_s$  determined by the Algorithm 3 be denoted by a user-defined parameter  $\mathbb{P} \in [0 : 1]$ , such that  $\mathbb{P} = 0$  and  $\mathbb{P} = 1$  represent zero and complete confidence, respectively.
4. generate a random number  $\mathcal{R} \sim U(0, 1)$  and if:
  - (a)  $\mathcal{R} < \mathbb{P}$ , then the probabilistic objective set, namely,  $\mathcal{F}_p$  is given by  $\mathcal{F}_p = \mathcal{F}_s$ .
  - (b)  $\mathcal{R} \geq \mathbb{P}$ , let  $\varphi$  denote the fraction of the redundant space (represented by  $\mathcal{F}_{redn}/\mathcal{W}_{redn}$ ) that needs to be accounted for (in addition to the *essential* space represented by  $\mathcal{F}_s/\mathcal{W}_s$ ) towards the formation of  $\mathcal{F}_p$ . Let  $\varphi$  be given by Equation 6.2, following which the composition of  $\mathcal{F}_p$  can be given by Equation 6.3.

$$\varphi = \frac{\mathcal{R} - \mathbb{P}}{1 - \mathbb{P}} \quad (6.2)$$

Notably, when: (i)  $\mathcal{R} = \mathbb{P}$ :  $\varphi = 0$  and  $\mathcal{F}_p = \mathcal{F}_s$ , (ii) as the value of  $\mathcal{R}$  grows in the range  $\mathbb{P} < \mathcal{R} < 1$ , the value of  $\varphi$  grows from zero to one, and more and more

of the redundant objectives will be incorporated<sup>3</sup> into  $\mathcal{F}_p$ , and (iii)  $\mathcal{R} = 1$ :  $\varphi = 1$  and  $\mathcal{F}_p = \mathcal{F}_0$ .

$$\left. \begin{aligned} \mathcal{F}_p &= \{f_{s1}, f_{s2}, \dots, f_{sJ}\}, \\ \text{where } J &\text{ is the minimum number determined by} \\ \sum_{j=1}^J w_{sj} &\geq \mathcal{W}_s + \varphi \mathcal{W}_{redn}. \end{aligned} \right\} \quad (6.3)$$

The determination of  $\mathcal{F}_p$  as mentioned above, is meant to be implemented during an MOEA run. In this way, every-time that a new random number  $\mathcal{R}$  is generated a new  $\mathcal{F}_p$  is determined accordingly. This ensures that the redundant objectives, besides the essential ones, are able take part during the MOEA search in a probabilistic manner. In the following section more details will be provided regarding the integration of  $\mathcal{F}_p$  during an MOEA run.

### 6.1.3 *Online* Implementation with Probabilistic Objective Sets

It has been stated that the *online* framework being proposed in this chapter aims to overcome the pitfall associated with the earlier proposed *offline* Framework 1, where an objective identified as redundant (by an application of the objective reduction algorithms), has no scope of being re-inducted into an MOEA's subsequent search. Towards this end, the probabilistic objective selection scheme introduces the scope for re-inducting an erroneously omitted objective. Furthermore, owing to the fact that the probabilistic scheme may help retain *all* the objectives (with different probabilities) at any given instant, it may also be useful in problems where the relationships between the objectives vary across the search space. To allow the probabilistic scheme to show its effect, the following approach is adopted:

1. beginning with a randomly initialised population and an initial objective set  $\mathcal{F}_0 = \{f_1, \dots, f_M\}$ , an MOEA is run for user-defined  $N_W$  generations to obtain a set of non-dominated solutions whose  $M$  objective vectors are stored in  $\mathcal{H}$  (set of objective vectors

---

<sup>3</sup>The fact that  $w_{sj}$  in Equation 6.3 represent the preference-weights for the objectives in decreasing order of importance, ensures that the objectives (including the redundant ones) with higher preference-weights are incorporated first into  $\mathcal{F}_p$ .

of the non-dominated solutions) that serves as the input data for the Algorithm 3. The *essential* objective set ( $\mathcal{F}_s$ ) and the preference-weights for all the objectives ( $w_i, i = 1, \dots, M$ ) are determined using the Algorithm 3, and this completes the first iteration of the *online* framework.

2. the subsequent iterations<sup>4</sup> involve:

(a) running the MOEA for another  $N_W$  generations, where *in each generation*:

- i. the probabilistic objective selection scheme (Section 6.1.2) is applied and  $\mathcal{F}_p$  is obtained and used, allowing the opportunity for an objective erroneously omitted from  $\mathcal{F}_s$  to be re-inducted.
- ii. the population of the previous generation becomes the current population on which the MOEA's selection and variation operators are applied with respect to the current  $\mathcal{F}_p$ .

(b) another application of the Algorithm 3 leading to a new  $\mathcal{F}_s$  and  $w_i$ s. Notably:

- i. the input for the Algorithm 3 is the  $M$  objective vectors of the non-dominated solutions in  $\mathcal{H}$  obtained after  $N_W$  generations of an MOEA.
- ii. the cardinality of  $\mathcal{F}_p$  corresponding to  $N_W^{th}$  generation of the MOEA may be  $m \leq M$ , i.e.,  $|\mathcal{F}_p| = m$ . In the case when  $m < M$ ,  $\mathcal{H}$  is composed by computing the missing  $M - m$  objective vectors corresponding to  $\mathcal{F}_0 \setminus \mathcal{F}_p$  based on the variable vectors in  $\mathcal{H}$ . This step ensures that  $\mathcal{H}$  is comprised of  $M$  objective vectors.

The above approach has been formalised and presented as Framework 4. In that, the number of iterations of the framework, are controlled by the user defined specifications.

---

<sup>4</sup>This is in contrast to the *offline* objective reduction by Framework 1, where the second iteration involves an MOEA run with just  $\mathcal{F}_s$  for another  $N_W$  generations followed by another application of the Algorithm 3 leading to a new  $\mathcal{F}_s$  and  $w_i$ s.

**Framework 4:** *Online objective reduction framework***Input:** $N_{gmax}$ : Maximum number of generations that the MOEA is allowed to run. $N_W$ : Number of generations per iteration of the framework. $\mathbb{P}$ : Probability of selecting the essential objective set  $\mathcal{F}_s$ .**1 begin****2**     $t = 0$ ,  $\mathcal{F}_t = \{f_1, \dots, f_M\}$  and  $\mathcal{H}$  (set of objective vectors of the non-dominated solutions).**3**    Generate a randomly initialised population corresponding to  $\mathcal{F}_t$  and store the population in  $\mathcal{H}$ .**4**    Beginning with  $\mathcal{H}$ , run the MOEA for  $N_W$  generations corresponding to the probabilistically determined objective set  $\mathcal{F}_p$  in each generation; and store the final population after  $N_W$  generations in  $\mathcal{H}$ . During the MOEA run:**5**    **if**  $t = 0$  **then****6**        $\mathcal{F}_p = \mathcal{F}_t$ .**7**    **else****8**       **for** *each generation* **do****9**          Generate a random number  $\mathcal{R} \sim U(0, 1)$ .**10**          **if**  $\mathcal{R} < \mathbb{P}$  **then****11**             The objective set selected is given by  $\mathcal{F}_p = \mathcal{F}_s$ .**12**          **else****13**             Determine  $\varphi$  (Equations 6.1 and 6.2).**14**             Determine  $\mathcal{F}_p$  (Equation 6.3).**15**          **end****16**       **end****17**    **end****18**    Apply Algorithm 3 to determine the *essential* objective set ( $\mathcal{F}_s$ ) and the preference-weights ( $w_i$ ), where the input for Algorithm 3, namely,  $\mathcal{H}$  is composed as below:**19**    **if**  $t = 0$  **then****20**       The  $M$  objective vectors of  $\mathcal{F}_t$  constitute  $\mathcal{H}$ .**21**    **end****22**    **else****23**       Let the cardinality of  $\mathcal{F}_p$  be  $m$ , i.e.,  $|\mathcal{F}_p| = m$ .**24**       Compose  $\mathcal{H}$  by computing the missing  $M - m$  objective vectors corresponding to  $\mathcal{F}_0 \setminus \mathcal{F}_p$  based on the variable vectors in  $\mathcal{H}$ .**25**    **end****26**    **if**  $\mathcal{F}_s = \mathcal{F}_t$  or  $N_W \times (t + 1) \geq N_{gmax}$  **then****27**       Stop and declare  $\mathcal{F}_s$  as the essential objective set.**28**    **else****29**       Set  $t = t + 1$ ,  $\mathcal{F}_t = \mathcal{F}_s$  and go to step 4.**30**    **end****31 end**

## 6.2 Experimental Settings and Quality Indicators

The proposed *online* framework relies on the use of an MOEA and an objective reduction algorithm. Here, NL-MVU-PCA is chosen as the objective reduction algorithm with parameters  $\theta = 0.997$  and  $q = M - 1$ , as suggested in Chapter 4. Furthermore,  $\epsilon$ -MOEA [34], a steady-state MOEA based on the  $\epsilon$ -dominance concept [35] is chosen as the underlying MOEA with the following parameter settings have been selected: (i) population size of 200; (ii) the probability of crossover and mutation as 0.9 and 0.1, respectively; (iii) the distribution index for crossover and mutation as 5 and 20, respectively; and (iv)  $\epsilon = 0.3$ . Note that the same parameters have been selected to conduct runs in Chapters 4 and 5.

For comparing the proposed *online* framework with the *offline* framework (Framework 1), both frameworks run  $\epsilon$ -MOEA with the above parameters and the number of generations per iteration is set to 2000 for all runs, leading to  $N_W = 2000$  for *online* and  $N_g = 2000$  for *offline*. For the *online* framework: (i) the maximum number of generations is set to  $N_{gmax} = 20000$ , implying a maximum of 10 iterations for the framework as long as the  $\mathcal{F}_s$  in two consecutive iterations continue to be different, and (ii)  $\mathbb{P}$ , the probability of picking  $\mathcal{F}_p = \mathcal{F}_s$  is fixed as  $0.682^5$ . Furthermore and for brevity, a non-dominated solution set obtained by the *online* and *offline* frameworks is denoted by  $\mathcal{N}_{\epsilon-p-On}$  and  $\mathcal{N}_{\epsilon-Off}$ , respectively.

The quality indicators for the POF-approximation obtained by the non-dominated solutions include  $g$  (a parameter in test problems) for convergence and  $I_s$  for diversity<sup>6</sup>.

## 6.3 Demonstration of the Framework 4 on DTLZ5(2,5)

This section demonstrates the working of the proposed *online* framework (Framework 4) on the DTLZ5(2,5) problem. This problem is chosen because it offers an opportunity to

---

<sup>5</sup>This is simply chosen in analogy with the *one-sigma* rule, where for a Gaussian distribution 68.2% of the values lie within one standard deviation of the mean.

<sup>6</sup>More details on quality indicators  $g$  and  $I_s$  are given in Section 4.4.

Table 6.1: Application of NL-MVU-PCA to DTLZ5(2, 5), corresponding to  $N_W = 2000$ .

Step	Section	Outcome
Eigenvalue analysis	4.1.3	$\mathcal{F}_e = \{f_1, f_2, f_3, f_4, f_5\}$
RCM analysis	4.1.4	$\{f_1, f_2, f_3, f_4\}$ and $\{f_5\}$
Selection scheme	4.1.5	$\mathcal{F}_s = \{f_4, f_5\}$ $\mathcal{E} = 0.000327$
Objective preference-weights	5.2.2	$w_1 = 0.000060$ $w_2 = 0.000069$ $w_3 = 0.000222$ $w_4 = 0.207757$ $w_5 = 0.791889$

highlight both the core components of the framework, namely objective reduction and probabilistic objective selection scheme. In that: (i) being a five-objective problem with a two dimensional POF, it offers a reasonable scope for objective reduction, and (ii) with only five objectives involved, it is quite manageable to discuss the probability of selection of different objectives based on their preference-weights. The implementation of framework based on  $\epsilon$ -MOEA as the underlying MOEA; NL-MVU-PCA as the underlying objective reduction algorithm, and with  $N_W = 2000$ , is presented below.

### 6.3.1 Objective Reduction

The application of NL-MVU-PCA leads to the results summarised in Table 6.1. In that:

1. the eigenvalue analysis could not omit any of the objectives, leading to  $\mathcal{F}_e = \{f_1, \dots, f_5\}$ ;
2. the RCM analysis identified  $\{f_1, f_2, f_3, f_4\}$  as the set of identically correlated objectives;
3. the selection scheme picked  $f_4$  as the best representative among the identically correlated  $f_1$ – $f_4$ , implying  $\mathcal{F}_s = \{f_4, f_5\}$  and  $\mathcal{F}_{redn} = \{f_1, f_2, f_3\}$ ;
4. the error incurred if  $\mathcal{F}_{redn}$  were to be eliminated is  $\mathcal{E} = 0.000327$ , where its components are  $\mathcal{E}_1 = 0.000060$ ,  $\mathcal{E}_2 = 0.000069$ , and  $\mathcal{E}_3 = 0.000222$ . Furthermore, the errors associated with the *essential* objectives are given by  $\mathcal{E}_4 = 0.207757$  and  $\mathcal{E}_5 = 0.791889$ ;
5. the objective preference-weights (normalised errors) in decreasing order of importance

are given by  $w_{s1} = w_5 = 0.791889$ ;  $w_{s2} = w_4 = 0.207757$ ;  $w_{s3} = w_3 = 0.000222$ ;  $w_{s4} = w_2 = 0.000069$ ; and  $w_{s5} = w_1 = 0.000060$ .

### 6.3.2 Post-processing: Probabilistic objective selection scheme

Given the above, the composition of the probabilistic objective sets ( $\mathcal{F}_p$ ) is discussed below. The Framework 4 prescribes that: (i) for  $\mathcal{R} < \mathbb{P}$ :  $\mathcal{F}_p = \mathcal{F}_s$ , and (ii) for  $\mathcal{R} \geq \mathbb{P}$ : the composition of  $\mathcal{F}_p$  is governed by Equation 6.3. The implementation of the  $\mathcal{R} \geq \mathbb{P}$  case can be appreciated by observing the following:

1.  $\mathcal{W}_s = \sum_{j=1}^2 w_{sj} = w_{s1} + w_{s2} = w_5 + w_4 = 0.999649$ .
2.  $\mathcal{W}_{redn} = w_{s3} + w_{s4} + w_{s5} = w_3 + w_2 + w_1 = 0.000351$ .
3.  $\sum_{j=1}^3 w_{sj} = w_{s1} + w_{s2} + w_{s3} = w_5 + w_4 + w_3 = 0.999871$ .
4.  $\sum_{j=1}^4 w_{sj} = w_{s1} + w_{s2} + w_{s3} + w_{s4} = w_5 + w_4 + w_3 + w_2 = 0.999940$ .
5.  $\sum_{j=1}^5 w_{sj} = w_{s1} + w_{s2} + w_{s3} + w_{s4} + w_{s5} = w_5 + w_4 + w_3 + w_2 + w_1 = 1.0$ .
6. Equation 6.3 suggests that when  $\mathcal{R} \geq \mathbb{P}$ , the redundant objectives will enter  $\mathcal{F}_p$  in the order of  $f_3$ ,  $f_2$  and  $f_1$  (in conformance with their preference-weights). The cut-off values of  $\mathcal{R}$  governing the inclusion of  $f_3$ ,  $f_2$  and  $f_1$  into  $\mathcal{F}_p$ , as summarised in Figure 6.1, can be determined as follows:

(a)  $\sum_{j=1}^3 w_{sj} = \sum_{j=1}^2 w_{sj} + \varphi \mathcal{W}_{redn}$ , implying:

- $\varphi = 0.63247863$ ; and  $\mathcal{R} = \mathbb{P} + \varphi(1 - \mathbb{P}) = 0.8831282$  (Equation 6.2).
- $\mathcal{F}_p = \{f_5, f_4\} \cup \{f_3\} = \{f_3, f_4, f_5\}$  for  $0.682 < \mathcal{R} \leq 0.8831282$ .

(b)  $\sum_{j=1}^4 w_{sj} = \sum_{j=1}^3 w_{sj} + \varphi \mathcal{W}_{redn}$ , implying:

- $\varphi = 0.82905983$ ; and  $\mathcal{R} = \mathbb{P} + \varphi(1 - \mathbb{P}) = 0.94564103$  (Equation 6.2).
- $\mathcal{F}_p = \{f_5, f_4\} \cup \{f_3, f_2\} = \{f_2, f_3, f_4, f_5\}$  for  $0.8831282 < \mathcal{R} \leq 0.94564103$ .



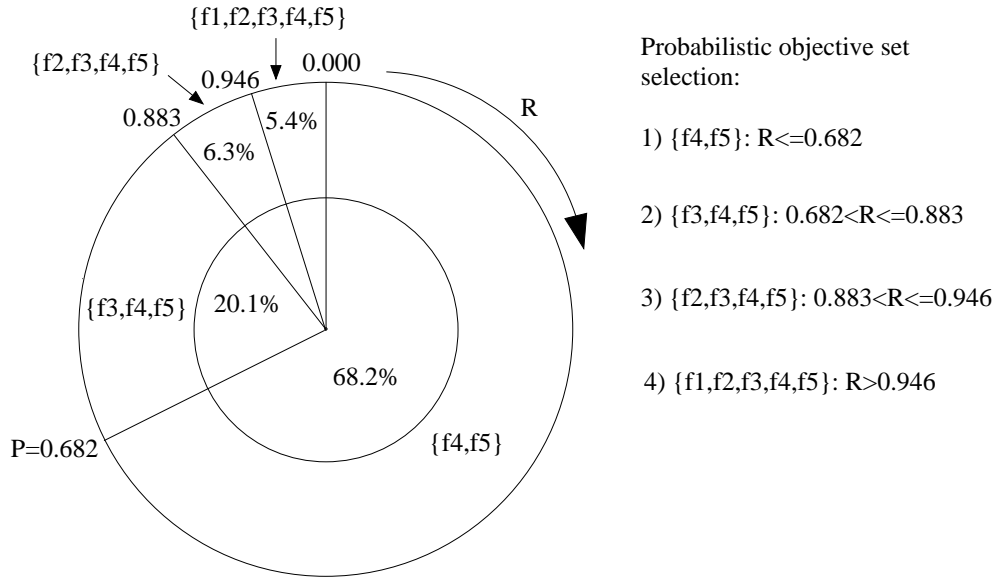


Figure 6.1: Different combinations of  $\mathcal{R}$  and  $\mathcal{F}_p$  for the DTLZ5(2, 5).

(c)  $\sum_{j=1}^5 w_{sj} = \sum_{j=1}^2 w_{sj} + \varphi \mathcal{W}_{redn}$ , implying:

- $\varphi = 1.0$ ; and  $\mathcal{R} = \mathbb{P} + \varphi(1 - \mathbb{P}) = 1.0$  (Equation 6.2).
- $\mathcal{F}_p = \{f_5, f_4\} \cup \{f_3, f_2, f_1\} = \{f_1, f_2, f_3, f_4, f_5\}$  for  $0.94564103 < \mathcal{R} \leq 1.0$ .

### 6.3.3 Online Implementation with Probabilistic Objective Sets

This section illustrates the performance of the Framework 4 on the DTLZ5(2, 5) problem, for which the objective reduction and probabilistic objective selection scheme has been discussed above. While the issue of how the Framework 4 can overcome the limitations of the *offline* Framework 1 in the wake of erroneous  $\mathcal{F}_s$  and/or  $w_i$ s (obtained by Algorithm 3) is deferred until the next section, Figure 6.2 reveals the drastic improvement in the performance (assessable through the  $g$  and  $I_s$  measures) of  $\epsilon$ -MOEA after  $N_W = 2000$  at which NL-MVU-PCA is applied, and  $\mathcal{F}_s$  and  $w_i$ s are obtained; and *after* which the  $\mathcal{F}_p$  is used in each generation.

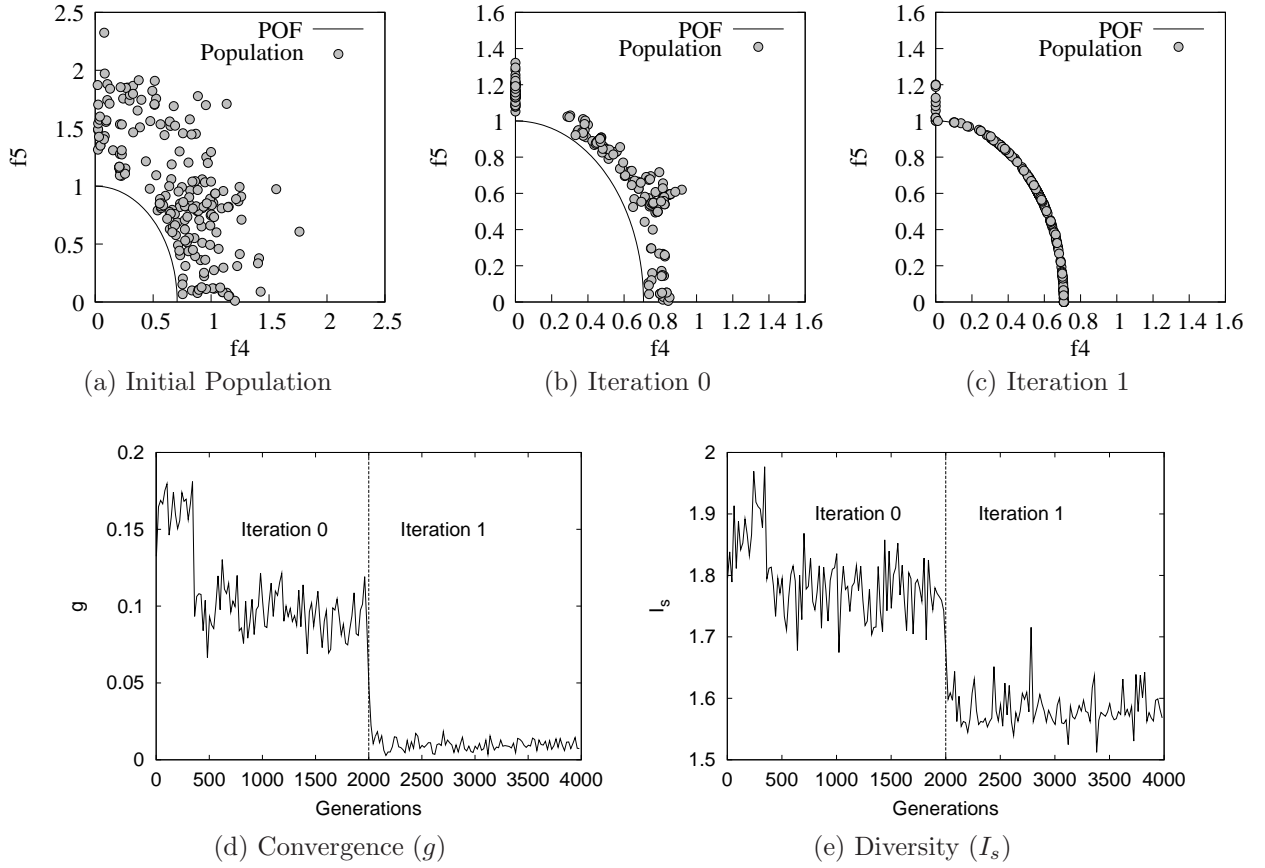


Figure 6.2: Application of the *online* Framework 4 to DTLZ5(2,5) (one run): Figure 6.2a depicts the random initial population vis-à-vis the true POF. Figures 6.2b and 6.2c presents the POF-approximation at  $N_W = 2000$  (0th iteration) and  $N_W = 4000$  (1st iteration), respectively. Figures 6.2d and 6.2e show the evolution of the quality indicators along the generations, namely,  $g$  for convergence and  $I_s$  for diversity. The dotted vertical line highlights  $N_W = 2000$  which indicates the application of the online objective reduction.

## 6.4 Handling of Erroneous Preference Information: *Online* Framework 4 versus *Offline* Framework 1

This section examines the ability of the *online* Framework 4 to handle erroneous preference information ( $\mathcal{F}_s$  and  $w_i$ s), vis-à-vis that of the *offline* Framework 1. Towards this, the same-problem (DTLZ5(2, 5)), underlying MOEA ( $\epsilon$ -MOEA), underlying objective reduction algorithm (NL-MVU-PCA), and the same parameter settings ( $N_W = 2000$  for *online*, and  $N_g = 2000$  for *offline*) as stated/used in previous sections, are retained.

Notably, in Section 6.3.1, application of NL-MVU-PCA on DTLZ5(2, 5) led to:

- $\mathcal{F}_s = \{f_4, f_5\}$  and  $\mathcal{F}_{redn} = \{f_1, f_2, f_3\}$ .
- $w_5 = 0.791891$ ,  $w_4 = 0.207757$ ,  $w_3 = 0.000222$ ,  $w_2 = 0.000069$  and  $w_1 = 0.000060$ .
- $\mathcal{W}_s = 0.999649$  and  $\mathcal{W}_{redn} = 0.000351$ .

To examine the performance with erroneous preference information, the above deductions by NL-MVU-PCA for DTLZ5(2, 5), are being overwritten in this section, by the following:

- $\mathcal{F}_s = \{f_1, f_2, f_3\}$ , even though  $f_1$ – $f_3$  are non-conflicting among themselves, in which case: (i)  $\mathcal{F}_s$  basically represents a single objective, and (ii) any MOEA population should collapse to a single solution.
- $w_1 = 0.40$ ,  $w_2 = 0.30$ ,  $w_3 = 0.20$ ,  $w_5 = 0.07$  and  $w_4 = 0.03$ .
- the above fixations of  $\mathcal{F}_s$  and  $w_i$ s, lead to the interplay of the random number  $\mathcal{R}$  and probabilistic objective sets  $\mathcal{F}_p$ , as summarised in Figure 6.3. In that:
  - $\mathcal{F}_p = \{f_1, f_2, f_3\}$  for  $0.0 < \mathcal{R} \leq 0.682$ .
  - $\mathcal{F}_p = \{f_1, f_2, f_3, f_5\}$  for  $0.682 < \mathcal{R} \leq 0.905$ .
  - $\mathcal{F}_p = \{f_1, f_2, f_3, f_4, f_5\}$  for  $0.905 < \mathcal{R} \leq 1.0$ .

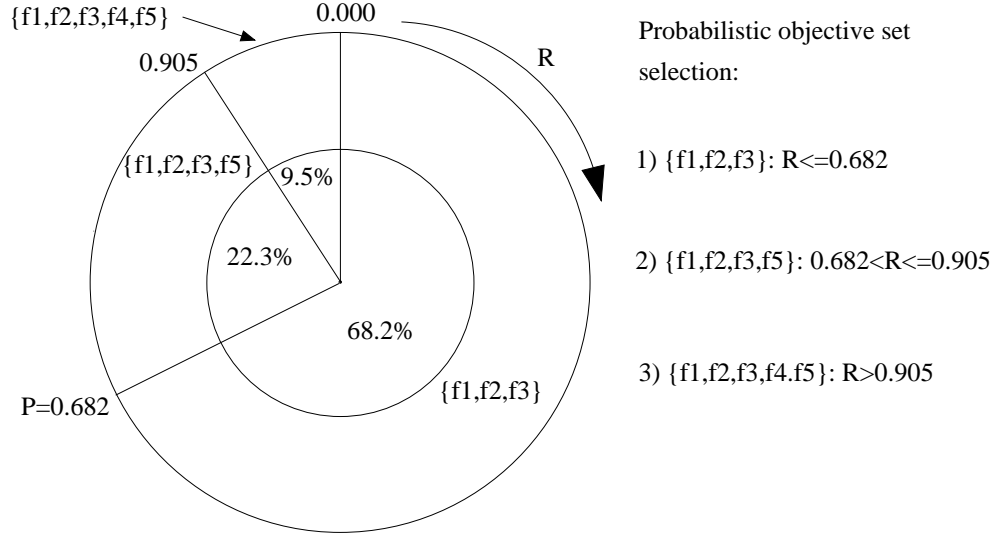


Figure 6.3: The interplay of the random number  $\mathcal{R}$  and probabilistic objective sets  $\mathcal{F}_p$ , corresponding to erroneous preference information.

Table 6.2: DTLZ5(2, 5): Performance of *online* Framework 4 versus *offline* Framework 1, while handling erroneous preference information

Here,  $N_W = 2000$  and  $N_g = 2000$  for *online* and *offline* frameworks, respectively, is used

(a) <i>Online</i> Framework							(b) <i>Offline</i> Framework	
Iteration	Objective preference-weights					$\mathcal{F}_s$	Iteration	$\mathcal{F}_s$
	$w_1$	$w_2$	$w_3$	$w_4$	$w_5$			
0	—	—	—	—	—	$\mathcal{F}_0$	0	$\mathcal{F}_0$
1	0.4000000	0.3000000	0.2000000	0.0300000	0.0700000	$\{f_1, f_2, f_3\}$	1	$\{f_1, f_2, f_3\}$
2	0.0162462	0.0578522	0.0002655	0.0000016	0.9256345	$\{f_2, f_5\}$	2	$\{f_1, f_2, f_3\}$
3	0.0000003	0.0000010	0.0000009	0.2494609	0.7505369	$\{f_4, f_5\}$		
4	0.0000025	0.0000005	0.0000032	0.1514993	0.8484946	$\{f_4, f_5\}$		

It can be seen from Table 6.2 that for the zeroth iteration, objective set for both the *online* Framework 4 and *offline* Framework 1 is  $\mathcal{F}_0 = \{f_1, \dots, f_5\}$ . It is just at the beginning of the first iteration (the generation at which objective reduction is to be applied:  $N_W = N_g = 2000$ ) that the erroneous preference information is imposed in the form of  $\mathcal{F}_s = \{f_1, f_2, f_3\}$  and  $\mathcal{W} = \{0.40, 0.30, 0.20, 0.03, 0.07\}$ , beyond which:

- the *offline* Framework 1 fails to alter the erroneous  $\mathcal{F}_s$  even after the second iteration. Given that, the same  $\mathcal{F}_s$  is retained in two successive iterations, namely, iteration 1 and 2, the framework is terminated.
- the *online* Framework 4 on the other hand, alters the  $\mathcal{F}_s$  in the second iteration to  $\mathcal{F}_s = \{f_2, f_5\}$ ; and to  $\mathcal{F}_s = \{f_4, f_5\}$  in the third iteration. As the same  $\mathcal{F}_s$  is retained in the fourth iteration, the framework is terminated.

Notably, while the the *offline* Framework 1 fails to recover the truly *essential* objective set  $\{f_4, f_5\}$  beginning with an erroneous  $\{f_1, f_2, f_3\}$ , the probabilistic objective selection scheme embedded in the *online* Framework 4 seems to serve as a self-correcting mechanism that enables the accurate identification of  $\mathcal{F}_s = \{f_4, f_5\}$ . More detailed observations and inferences about the performance of both the frameworks, can be made with respect to the Figure 6.4, as follows:

- ideally the *offline* Framework 1 working with the set of non-conflicting objectives, namely,  $\{f_1, f_2, f_3\}$ , should collapse to a single solution. However, given the fact that this is a many-objective problem ( $M = 5$ ), where the presence of too many objectives detracts the convergence, the framework could at-best reduce the entire population to an ensemble of closely located solutions as seen in Figure 6.4a. This pattern justifies the extremely poor values (deviating from the ideal value of  $I_s = 1$ ) of the diversity indicator  $I_s$  in Figure 6.4f and also explains the misleading good values (close to the ideal value of  $g = 0$ ) for the convergence indicator  $g$  in Figure 6.4e.

- Figure 6.4b suggests that the *online* Framework 4, even while working with  $\mathcal{F}_s = \{f_1, f_2, f_3\}$  in the first iteration, is able to include  $f_4$  and  $f_5$  in  $\mathcal{F}_p$  despite their low preference-weights. The inclusion of  $f_4$  and  $f_5$  in  $\mathcal{F}_p$ , and hence in the  $\epsilon$ -MOEA search, allows the subsequent applications of the NL-MVU-PCA to detect the true relationships between the objectives with increasing accuracy ( $\mathcal{F}_s = \{f_2, f_5\}$  in the second iteration, and  $\mathcal{F}_s = \{f_4, f_5\}$  in the third iteration). This argument is supported by:
  - the visible improvement in the POF-approximation over subsequent iterations, in Figures 6.4c and 6.4d.
  - the tendency of the convergence and diversity indicators, namely,  $g$  and  $I_s$ , to acquire their ideal values of 0 and 1, respectively, as in Figures 6.4g and 6.4h.

## 6.5 Broader Set of Experimental Results

This section presents the experimental results for a broader set of redundant test problems, including: (i) DTLZ5( $I, M$ ) instances, where  $I$  is scaled up to 7 and  $M$  up to 50; and (ii) WFG3( $M$ ) instances, where  $M$  is scaled up to 25. For each problem, both the *online* Framework 4 and *offline* Framework 1 are applied for 20 different random seeds (required by the underlying MOEA, namely,  $\epsilon$ -MOEA) and their performance is compared in terms of: (i) the frequency of their success—the number of times out of the 20, that these frameworks are able to accurately identify the dimension and composition of the true *essential* objective set ( $\mathcal{F}_T$ ), and (ii) the number of iterations required. The same parameter settings as in the previous chapters/sections are retained, notable among which are  $N_W = 2000$  and  $N_g = 2000$ , as they imply similar computational effort for both the frameworks.

The results corresponding to the above test-suite are presented in Table 6.3. In that:

- for the DTLZ5(2,  $M$ ) and DTLZ5(3,  $M$ ) problems with varying  $M$ : the accuracy of both the frameworks can be seen to be identical in terms of the accuracy and required

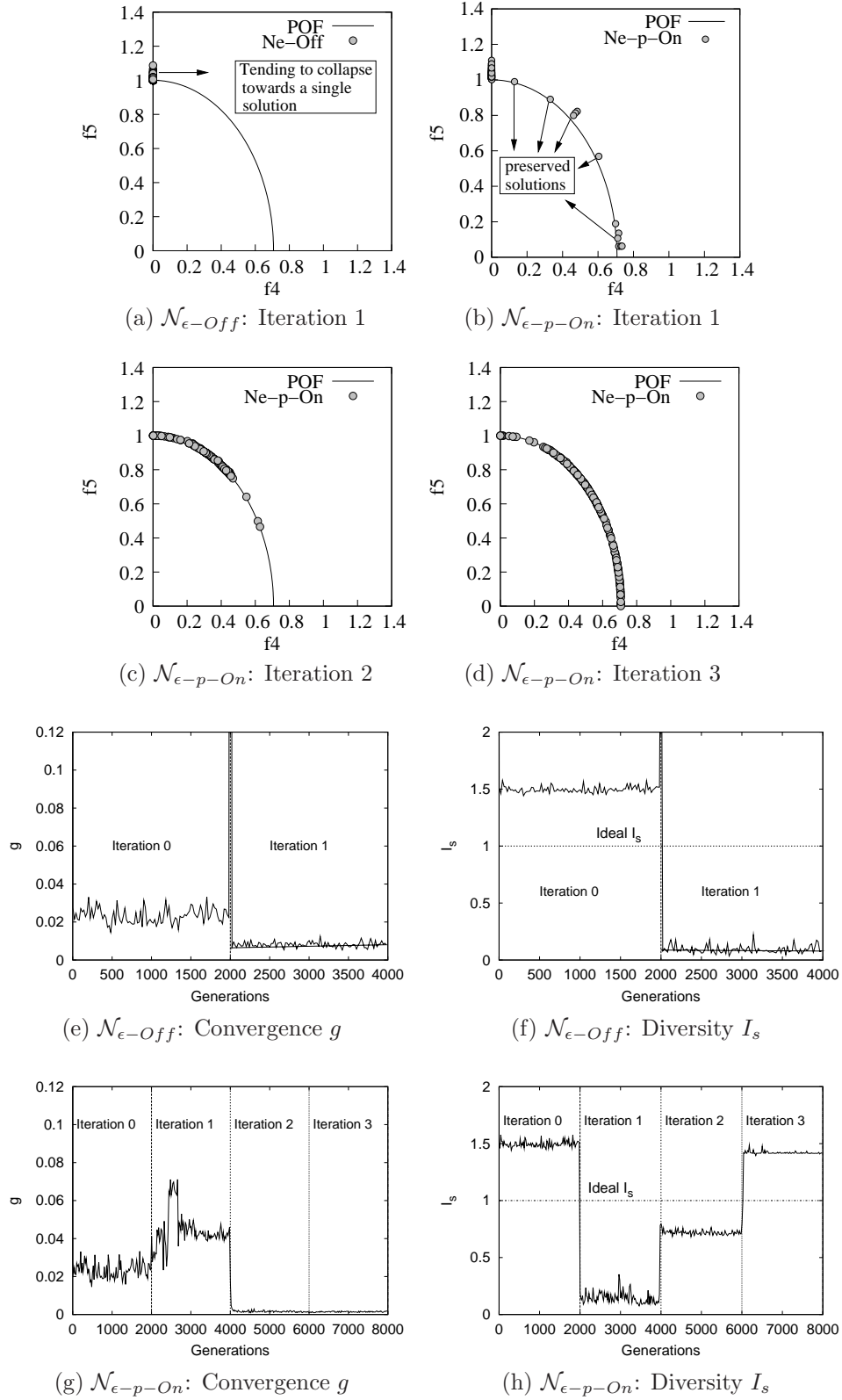


Figure 6.4: DTLZ5(2, 5): POE-approximation obtained by the *online* Framework 4 and *offline* Framework 1, while handling erroneous preference information. Here,  $N_W = 2000$  and  $N_g = 2000$  for *online* and *offline* frameworks, respectively, is used.

computational effort. This could be attributed to the fact that the DTLZ5( $I, M$ ) problems corresponding to  $I = 2$  and 3 are relatively simpler compared to those with  $I = 5$  or 7. In such cases, as the Table 6.4 suggests, the POF-approximation by  $\mathcal{N}_{\epsilon\text{-}Off}$  is good enough (notably,  $\mathcal{N}_{\epsilon\text{-}p\text{-}On}$  is better) to allow the Framework 1 to solve accurately.

- for the DTLZ5(5,  $M$ ) and DTLZ5(7,  $M$ ) problems with varying  $M$ : the accuracy of the Framework 4 is higher than that of the Framework 1. This could be attributed to relatively more difficult nature of these problems, and hence, poorer POF-approximation by  $\mathcal{N}_{\epsilon\text{-}Off}$  compared to  $\mathcal{N}_{\epsilon\text{-}p\text{-}On}$ , as evident in Table 6.4. Notably, in some instances, Framework 4 requires more iterations (as shown in Table 6.3) but this could be explained by the fact that the probabilistic objective selection scheme allows for some redundant objectives to participate during the MOEA search, and in effect delays accurate results.
- the same trend of higher accuracy by the Framework 4 (compared to the Framework 1) at the cost of higher computational effort is evident in the case of WFG3( $M$ ) problems.

## 6.6 Summary

This chapter proposed an *online* objective reduction framework that employs a probabilistic selection scheme for dealing with MaOPs. This framework aims to address the limitations of the *offline* framework in terms of its inability to: (i) re-induct an *essential* objective that may erroneously be omitted by an objective reduction algorithm, and/or (ii) adapt to the varying relationships between the objectives, across the search space. With reference to a sample problem, it has been discussed in detail as to how the probabilistic retention of all the objectives (including those inferred as redundant) serves as a self-correcting mechanism for the proposed *online* framework. Experiments on a broad set of problems has revealed



Table 6.3: Performance comparison between the *online* Framework 4 and *offline* Framework 1 with  $\epsilon$ -MOEA as the underlying MOEA (data:  $\mathcal{N}_\epsilon$ ) and NL-MVU-PCA as the underlying objective reduction algorithm

The entries below  $I$  and  $\mathcal{F}_T$  labels indicate frequency of success in 20 different runs. The entries below the  $N_{obj}$  label are formatted as  $\mu \pm \sigma$ , where  $\mu$  and  $\sigma$  represent the mean and standard deviation, respectively, of the number of essential objectives identified, in 20 runs. The entries below Iterations label report cases requiring multiple iterations for accurate results, where  $aR(bI)$  implies that  $a$  runs required  $b$  iterations each; while  $(—)$  indicate cases requiring only one iteration

Test Problem	$\mathcal{N}_{\epsilon-Off}$ (Framework 1 on $\mathcal{N}_\epsilon$ )				$\mathcal{N}_{\epsilon-p-On}$ (Framework 4 on $\mathcal{N}_\epsilon$ )			
	$I$	$\mathcal{F}_T$	$N_{obj}$	Iterations	$I$	$\mathcal{F}_T$	$N_{obj}$	Iterations
DTLZ5(2, 05)	20	20	$02.0 \pm 0.0$	—	20	20	$02.0 \pm 0.0$	—
DTLZ5(2, 20)	20	20	$02.0 \pm 0.0$	—	20	20	$02.0 \pm 0.0$	—
DTLZ5(2, 50)	20	20	$02.0 \pm 0.0$	—	20	20	$02.0 \pm 0.0$	—
DTLZ5(3, 05)	20	20	$03.0 \pm 0.0$	—	20	20	$03.0 \pm 0.0$	—
DTLZ5(3, 20)	20	20	$03.0 \pm 0.0$	—	20	20	$03.0 \pm 0.0$	—
DTLZ5(5, 10)	19	19	$05.0 \pm 0.2$	3R(2I)	20	20	$05.0 \pm 0.0$	2R(2I) 2R(3I)
DTLZ5(5, 20)	19	19	$05.0 \pm 0.2$	8R(2I)	20	20	$05.0 \pm 0.0$	4R(2I) 2R(3I) 1R(4I)
DTLZ5(7, 10)	19	19	$07.0 \pm 0.2$	7R(2I)	20	20	$07.0 \pm 0.0$	4R(2I) 1R(3I) 1R(4I) 1R(5I)
DTLZ5(7, 20)	16	16	$07.2 \pm 0.4$	9R(2I) 1R(4I)	20	20	$07.0 \pm 0.0$	4R(2I) 3R(3I) 3R(4I) 2R(6I) 1R(8I)
WFG3(05)	20	20	$02.0 \pm 0.0$	—	20	20	$02.0 \pm 0.0$	—
WFG3(15)	15	15	$02.2 \pm 0.4$	6R(2I) 3R(3I)	15	15	$02.7 \pm 1.3$	2R(2I) 1R(3I) 5R(4I) 3R(5I) 1R(8I) 1R(9I)
WFG3(25)	09	09	$02.7 \pm 0.7$	5R(2I)	15	15	$02.6 \pm 1.2$	2R(2I) 1R(3I) 1R(4I) 3R(5I) 1R(6I) 1R(7I) 2R(8I) 2R(9I)

Table 6.4: On POF-approximation by  $\mathcal{N}_{\epsilon-p-On}$  and  $\mathcal{N}_{\epsilon-Off}$ , accessed by convergence ( $g$ ) and diversity ( $I_s$ ) measures

The measures $g$ and $I_s$ are averaged over 20 runs				
Test Problems	$\mathcal{N}_{\epsilon-p-On}$		$\mathcal{N}_{\epsilon-Off}$	
	$g$	$I_s$	$g$	$I_s$
DTLZ5(2, 05)	$0.001165 \pm 0.000730$	$1.424381 \pm 0.035902$	$0.018435 \pm 0.011791$	$2.190684 \pm 0.542994$
DTLZ5(2, 20)	$0.000903 \pm 0.000158$	$1.415556 \pm 0.001684$	$0.024209 \pm 0.020711$	$2.203498 \pm 0.505308$
DTLZ5(2, 50)	$0.000993 \pm 0.000196$	$1.416205 \pm 0.002169$	$0.029989 \pm 0.018604$	$2.244674 \pm 0.314031$
DTLZ5(3, 05)	$0.013709 \pm 0.004215$	$1.249509 \pm 0.039814$	$0.037362 \pm 0.007345$	$1.334061 \pm 0.038721$
DTLZ5(3, 20)	$0.012866 \pm 0.005509$	$1.243917 \pm 0.051771$	$0.039143 \pm 0.010904$	$1.351592 \pm 0.054949$
DTLZ5(5, 10)	$0.040170 \pm 0.011232$	$1.173417 \pm 0.035820$	$0.688565 \pm 0.044774$	$2.022492 \pm 0.108477$
DTLZ5(5, 20)	$0.044578 \pm 0.013947$	$1.186967 \pm 0.052760$	$0.109226 \pm 0.020977$	$1.344631 \pm 0.031619$
DTLZ5(7, 10)	$0.048867 \pm 0.012230$	$1.095047 \pm 0.048257$	$0.153006 \pm 0.022800$	$1.300638 \pm 0.032023$
DTLZ5(7, 20)	$0.051428 \pm 0.012024$	$1.100106 \pm 0.038208$	$0.142368 \pm 0.024472$	$1.270626 \pm 0.054090$

that the enhanced accuracy of the *online* framework may come at the cost of higher computational effort. This is owing to the fact that on the one hand the probabilistic inclusion of the objectives inferred as redundant promises enhanced accuracy in the face of erroneous inferences on redundancy, on the other hand it makes the MOEA search more complex and hence, delays the capturing of true relationships between the objectives. In terms of the state-of-the-art, while the *online* implementation of the algorithms based on DRP [2] and UFS [3] has focussed on the non-redundant problems, this is the first study that demonstrates the benefits of *online* objective reduction on redundant test problems.

## Chapter 7

# *Timing* the Decision Support

In the previous chapters, the *offline* and *online* implementations of the objective reduction framework have been proposed. In both cases, the framework conducts an MOEA run for a pre-specified number of generations ( $N_g$ ) and the final population (from the last generation) serves as the input data for the objective reduction algorithms. However, there is neither any criterion in the existing literature on objective reduction available on how to determine  $N_g$ , nor is there any general termination criterion available in the MOEA domain that could be adapted to determine  $N_g$  in the context of objective reduction for MaOPs.

This thesis attempts to address this fundamental issue by recognising that the proposed framework relies on a good POF-representation, and a *sufficient* condition for it is a good POF-approximation. Hence, the underlying MOEA could be terminated and the framework be applied, if any of the following conditions could be ascertained:

- (a) that, the MOEA has converged to the POF, i.e., a good POF-approximation has been obtained;
- (b) that, the MOEA has stagnated, implying its inability to further improve the POF-approximation.

The rationale for the above argument lies in the fact that in either of the above cases, running the MOEA any further will offer no further improvement in the POF-approximation, despite

additional computational expense. In this background, this thesis proposes an entropy based *dissimilarity* measure, which determines the generation count ( $N_{gt}$ ) at which it could be ascertained that the *sufficiency* condition of a good POF-approximation has been met, or that it can not be met. Significantly, the computational complexity of the proposed *dissimilarity* measure is low (linear in nature), and the measure is generic enough to make inferences about the POF-approximation in the case of two- and three-objective problems. In the case of MaOPs, the measure is able to indicate the stagnation of NSGA-II, i.e., its inability to improve the quality of the POF-approximation any further.

## 7.1 Entropy and Relative Entropy

In information theory, *entropy* is a basic concept introduced by Shannon [166] and it is also known as *Shannon entropy*. It is generally used as a key measurement of information. The entropy measures the uncertainty associated with the prediction of the outcome of a random variable. Consider the probability distribution  $p(X)$  where  $X \in \mathbb{R}^M$  is a random vector. Entropy is defined as

$$\mathcal{H}(X) = - \int_{\mathbb{R}^M} p(X) \log p(X) dX, \quad (7.1)$$

or in discrete domain, that is  $X \in \mathbb{L}^M$  where  $\mathbb{L}$  is a discrete set, the entropy is defined as,

$$\mathcal{H}(X) = - \sum_{i=1}^T p(x_i) \log p(x_i), \quad (7.2)$$

where  $x_i$  is an element of the discrete space  $\mathbb{L}^M$ , and the cardinality of the space  $\mathbb{L}^M$  is  $T$ . The previous measurement only quantifies the information within the probability distribution  $p(x_i)$ . For comparing two different distributions a concept known as *relative entropy* (also known as Kullback–Leibler divergence [167]) quantifies how close a probability distribution  $p(x_i)$  is to a model (or candidate) distribution  $q(x_i)$ . It can be used as a *dissimilarity* measure

between two stochastic processes. This measurement can be expressed as

$$\mathcal{KL}(p||q) = - \int_{\mathbb{R}^M} p(X) \log \left\{ \frac{q(X)}{p(X)} \right\} dX, \quad (7.3)$$

or in discrete domain as

$$\mathcal{KL}(p||q) = - \sum_{i=1}^T p(x_i) \log \left\{ \frac{q(x_i)}{p(x_i)} \right\}. \quad (7.4)$$

Notably, the following characteristics hold true for  $\mathcal{KL}(p||q)$ :

1. It is always non-negative, i.e.,  $\mathcal{KL}(p||q) \geq 0$ .
2. It is not symmetric since  $\mathcal{KL}(p||q) \neq \mathcal{KL}(q||p)$ .
3. Only in a case when  $p(X) = q(X)$  then  $\mathcal{KL}(p||q) = \mathcal{KL}(q||p)$ .

## 7.2 Probability Distribution Estimation

To compute either entropy or the relative entropy as mentioned in the previous section, it is necessary to estimate the probability distribution, referred to as  $p(x_i)$  and  $q(x_i)$ . For this purpose, one needs to select a probability distribution estimation method, which can be either: (i) parametric, (ii) semi-parametric or (iii) non-parametric [168]. The pros and cons of each category are described as follows:

1. The parametric methods are characterised for having a specific probability distribution *functional form* that is determined by a small number of parameters whose values are captured from the dataset. In multivariate statistics a widely used parametric method is the multivariate Gaussian distribution which is governed by the mean vector and the covariance matrix. Due to those characteristics the distribution estimated by the Gaussian is unimodal. If the process that generated the data is multimodal, which is common on real-world datasets, the distribution will be poorly represented by a

Gaussian. As a consequence, the parametric methods are not flexible enough to model all kinds of distributions that may have been used to generate the data.

2. The semi-parametric methods are a combination of parametric and non-parametric methods. It is common to use them when it is intended to apply a parametric method but the *functional form* of the distribution is not known. However, the estimation of a large number (flexibility of the model increases with number of parameters) of parameters in the semi-parametric model is a time consuming process. Hence, if these parameters are to be estimated for each generation, it becomes an extremely time-demanding task. Besides, since they still rely on parametric methods, such as Gaussian, they are also governed by the same assumptions and hence they might not be flexible enough to model a real-world dataset distribution.
3. The most flexible probability distribution estimation methods are non-parametric since they make a few assumptions about the probability distribution *functional form*. For these methods the density estimation is driven entirely by the data. Some available methods in this category are: (i) multivariate histograms [169] and (ii) kernel density estimation [170, 171]. The multivariate histogram method is simple and fast, and hence, it is described in more detail below.

The multivariate histogram method partitions each dimension into a fixed number of intervals (bins) defined by an anchor point (in here, the origin) and  $M$  bin widths  $h_1, \dots, h_M$ , each corresponding to one dimension. In multidimensional space the partitions define a rectangular cell with hypervolume given by  $\prod_{j=1}^M h_j$ . Then, the probability distribution function associated with cell  $x_i$  is given by

$$p(x_i) = \frac{k(x_i)}{N}, \quad (7.5)$$

where  $N$  represents the total number of data points, and  $k(x_i)$  denotes the number of data points that exist in the cell  $x_i$ . The number of cells required to partition the search space

is a general problem associated with the multidimensional histogram method, as these can grow exponentially with the increase in the dimensions and in the worst case it can lead to  $n_{bin}^M$ , where  $n_{bin}$  is the number of bins per dimension. Notably, the number of cells (on which the smoothing of the probability distribution depends) is defined by the bin widths, in that, if the bin width reduces, the number of cells increase.

### 7.3 Multidimensional Histogram Algorithm for MOEA Populations

While the concept of multidimensional histogram has been introduced above, this section aims to utilise it in the context of MOEA populations.

To lay the foundation, let two data sets be given by  $P'$  and  $Q'$ , with  $P$  and  $Q$  representing their minimal sets, respectively; and  $p$  and  $q$  denoting their probability distributions, respectively. Furthermore, let the intersecting and the non-intersecting regions between  $P$  and  $Q$  be defined as follows:

1. Let  $J_I$  be a set of cells that represents the intersection region between sets  $P$  and  $Q$ , where each cell satisfy the conditions  $p(x_i) > 0$  and  $q(x_i) > 0$ . In this way, each cell from  $J_I$  is comprised by solutions from both sets,  $P$  and  $Q$ .
2. Let  $P_{NI}$  and  $Q_{NI}$  be two sets of cells that represent the non-intersection region for sets  $P$  and  $Q$ , respectively. The sets can be differentiated by the types of solutions existent in their cells as follows:
  - (a) In  $P_{NI}$  each cell satisfy the conditions  $p(x_i) > 0$  and  $q(x_i) = 0$ . This means that only solutions from  $P$  exist in each cell.
  - (b) In  $Q_{NI}$  each element satisfy the conditions  $p(x_i) = 0$  and  $q(x_i) > 0$ . This means that only solutions from  $Q$  exist in each cell.

Notably, the defined sets are disjoint since  $J_I \cap P_{NI} = \emptyset$ ,  $J_I \cap Q_{NI} = \emptyset$  and  $P_{NI} \cap Q_{NI} = \emptyset$ . Also, the total region in the search space occupied by population  $P'$  is given by  $J_I \cup P_{NI}$ , while for population  $Q'$  is given by  $J_I \cup Q_{NI}$ .

In the current context, where the aim is to identify when an MOEA population stabilises, let  $P$  and  $Q$  represent the MOEA population (objective vectors of the non-dominated solutions that are feasible) in two successive generations, with  $p$  and  $q$  representing their probability distributions, respectively. In the context of MOEAs, where  $M$  and  $N$  could be large, partitioning the objective space in a computationally efficient manner will pose a major challenge—an issue that is addressed below.

### 7.3.1 Computationally Efficient Data Structures

To counter the difficulty associated with partitioning of the objective space in a computationally efficient manner:

1. it is assumed that all the bins have the same width;
2. the use of the following data structures is proposed:
  - (a)  $C$ : vector that stores cells from regions  $J_I$  and  $P_{NI}$ ;
  - (b)  $C_q$ : vector that stores cells from region  $Q_{NI}$ .

To appreciate the effectiveness of these data structures, consider the two sample illustrations in Figure 7.1, where the two populations  $P$  and  $Q$  comprise of six solutions each; and the multidimensional histogram is set for five bins, i.e.,  $n_{bin} = 5$ . Since the proposed data structures account for only those cells where solutions exist:

- (i) Figure 7.1a: the total number of cells required to partition both populations is only 9 as opposed to 25. In that, the vector  $C$  stores 5 cells corresponding to  $J_I$  and  $P_{NI}$  (where  $|J_I| = 2$  and  $|P_{NI}| = 3$ ), and vector  $C_q$  stores 4 cells corresponding to  $Q_{NI}$  (implying  $|Q_{NI}| = 4$ ).



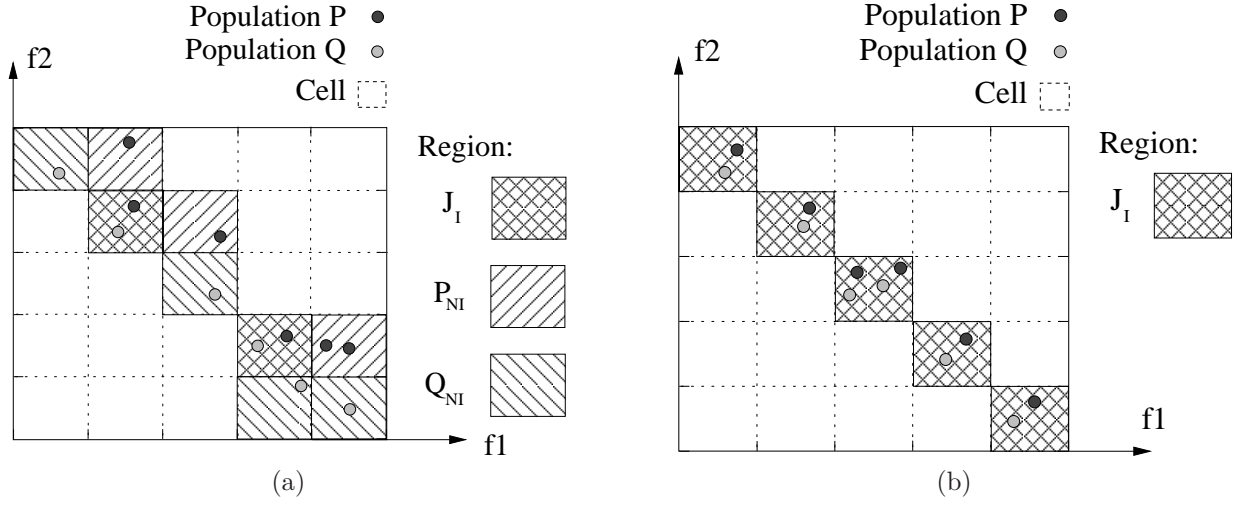


Figure 7.1: Effect of population dispersion in the number of cells and their corresponding region.

- (ii) Figure 7.1b: since only the intersection region exist, the number of required cells is given by  $|J_I| = 5$ , which leads to  $|C| = 5$  and  $|C_q| = 0$ .

Notably, given the choice of the data structures, the computational complexity reduces from  $n_{bin}^M$  to  $2N$ , i.e., independent of the number of objectives involved and only depends on the number of solutions.

Furthermore, in order to calculate the probability of finding a solution in a cell it is important to do it in respect to each population. Towards it, the use of the following data structures is proposed:

1.  $P_c$ : a vector that stores the number of solutions within each cell of  $C$  for population  $P$ . This is necessary for calculating the probability of finding a solution from population  $P$  in a cell from regions  $J_I$  and  $P_{NI}$ ;
2.  $Q_c$ : a vector that stores the number of solutions within each cell of  $C$  for population  $Q$ . This is necessary for calculating the probability of finding a solution from population  $Q$  in a cell from region  $J_I$ ;
3.  $Q_{cq}$ : a vector that stores the number of solutions within each cell of  $C_q$  for population  $Q$ .

This is necessary for calculating the probability of finding a solution from population  $Q$  in a cell from region  $Q_{NI}$ ;

### 7.3.2 Assignment of a Unique Identification Number to a Cell

Before a cell is stored inside any data structure, such as  $C$  or  $C_q$ , it needs a unique number that would identify it. Towards this, the cell identification procedure is as follows:

1. Let  $X_{i,j}$  define the value of the  $i^{th}$  solution in the  $j^{th}$  objective, where  $i = 1, \dots, N$  and  $j = 1, \dots, M$ . Also, let  $O_{max,j}$  and  $O_{min,j}$  define the maximum and minimum values, respectively, for objective  $j$  between all solutions.
2. A function named *GetCell\_id* returns a value ( $c$ ) which is used to identify the cell that a solution in objective space belongs to. In that, consider the following steps:
  - (a) map the solution into the range  $[0, 1]$  by:

$$\bar{X}_{i,j} = \frac{X_{i,j} - O_{min,j}}{O_{max,j} - O_{min,j}}, \text{ for } j = 1, \dots, M; \quad (7.6)$$

- (b) let a vector  $B = \left\{ \frac{0}{n_{bin}}, \frac{1}{n_{bin}}, \dots, \frac{n_{bin}}{n_{bin}} \right\}$  with size  $n_{bin} + 1$  define a set of intervals such that  $B_{k_j} \leq \bar{X}_{i,j} \leq B_{k_j+1}$  where  $k_j \in [0, \dots, n_{bin} - 1]$ ;
- (c) finally, a unique cell identification number is given by:

$$c = \sum_{j=1}^M k_j \times n_{bin}^{j-1}. \quad (7.7)$$

To appreciate the above, consider a sample illustration in Figure 7.2 for a two-objective case, with  $n_{bin} = 5$ . Three solutions are shown with coordinates: (i)  $A(0.36, 0.96)$ , (ii)  $B(0.78, 0.78)$  and (iii)  $C(0.71, 0.61)$ . The cell identification number for solution  $A$ , where  $k = \{1, 4\}$ , is given by  $c = (1 \times 5^0) + (4 \times 5^1) = 21$ . For solutions  $B$  and  $C$ , where  $k = \{3, 3\}$ ,

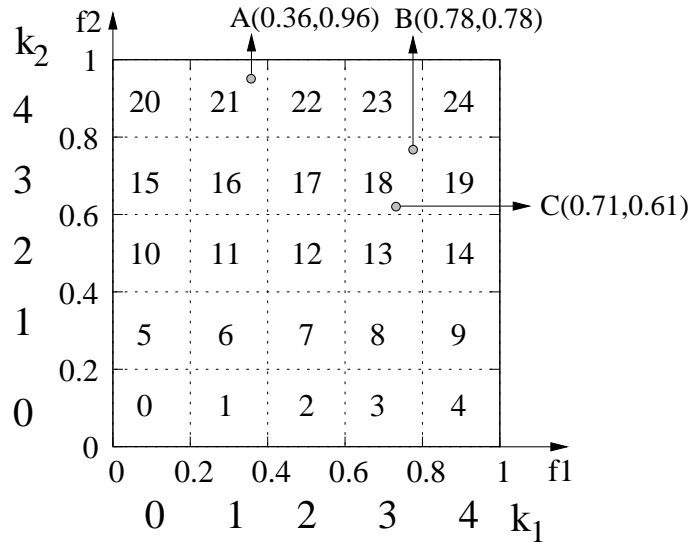


Figure 7.2: Unique cell identification number attributed to each cell by Equation 7.7, where  $n_{bin} = 5$  leads to 25 cells for a two-objective problem. Solution A is attributed the cell identification number 21, while solutions B and C are assigned 18.

the cell identification number is given by  $c = (3 \times 5^0) + (3 \times 5^1) = 18$ . Note that, solution B and C belong to the same cell, hence, their cell identification number is the same.

### 7.3.3 Multidimensional Histogram Algorithm: General Steps

In this section, the general steps of the multidimensional histogram algorithm (Algorithm 5) are summarised as follows:

1. Find a cell corresponding to each solution in  $P$ , by using *GetCell\_id* function (described in Section 7.3.2). If the cell already exists in vector  $C$ , increment by one the element in the vector  $P_c$  that corresponds to the identified cell. Otherwise, proceed as follows:
  - (a) add the cell to vector  $C$  to keep a track of all found cells for population  $P$ ;
  - (b) let vector  $P_c$  know that a solution was found by initialising the corresponding cell position with value 1;
  - (c) for the same position initialise vector  $Q_c$  with value 0 since no solution from population  $Q$  was found so far for that cell.

---

**Algorithm 5:** Multidimensional Histogram Algorithm for two MOEA populations.

---

**Input:** $P$ : feasible and non-dominated population corresponding to instant  $t$  $Q$ : feasible and non-dominated population corresponding to instant  $t + 1$  $n_{bin}$ : number of bins used to partition the search space equally among all objectives**Output:** $C$ : vector that stores cells from regions  $J_I$  and  $P_{NI}$  $C_q$ : vector that stores cells from region  $Q_{NI}$  $P_c$ : vector that stores the number of solutions within each cell of  $C$  for population  $P$  $Q_c$ : vector that stores the number of solutions within each cell of  $C$  for population  $Q$  $Q_{cq}$ : vector that stores the number of solutions within each cell of  $C_q$  for population  $Q$ 

```

1  begin
2      for each  $s$  in  $P$  do
3           $c \leftarrow GetCell\_id(s, n_{bin})$ 
4          if  $c$  exist in  $C$  then
5               $k = \{\text{index position of } c \text{ in vector } C\}$ 
6               $P_{c,k} = P_{c,k} + 1$                                 /* increment  $P_{c,k}$  by 1 */
7          else
8               $C = \{C, c\}$                                     /* concatenate  $c$  at the end of  $C$  */
9               $P_c = \{P_c, 1\}$                                 /* concatenate 1 at the end of  $P_c$  */
10              $Q_c = \{Q_c, 0\}$                                 /* concatenate 0 at the end of  $Q_c$  */
11         end
12     end
13     for each  $s$  in  $Q$  do
14          $c \leftarrow GetCell\_id(s, n_{bin})$ 
15         if  $c$  exist in  $C$  then
16              $k = \{\text{index position of } c \text{ in vector } C\}$ 
17              $Q_{c,k} = Q_{c,k} + 1$                                 /* increment  $Q_{c,k}$  by 1 */
18         else
19             if  $c$  exist in  $C_q$  then
20                  $k = \{\text{index position of } c \text{ in vector } C_q\}$ 
21                  $Q_{qc,k} = Q_{qc,k} + 1$                         /* increment  $Q_{qc,k}$  by 1 */
22             else
23                  $C_q = \{C_q, c\}$                             /* concatenate  $c$  at the end of  $C_q$  */
24                  $Q_{qc} = \{Q_{qc}, 1\}$                         /* concatenate 1 at the end of  $Q_{qc}$  */
25             end
26         end
27     end
28 end

```

---

In this step the procedure counted the number of solutions that fall into each cell from population  $P$ . Each cell is stored in vector  $C$  while the number of solutions in vector  $P_c$ . Also, the corresponding position in vector  $Q_c$  is initialised with value 0. Steps 2-12 in Algorithm 5.

2. Find a cell corresponding to each solution in  $Q$ , by using *GetCell\_id* function. If the cell already exists in vector  $C$ , increment the corresponding position in vector  $Q_c$ . Otherwise, proceed as follows:

- (a) If the cell exists in vector  $C_q$ , increment by one the element in the vector  $Q_{cq}$ .
- (b) Else: (i) add the cell to vector  $C_q$  to keep a track of all found cells for population  $Q$  that were not found for population  $P$  and (ii) initialise with value 1 the corresponding position in vector  $Q_{cq}$ .

In this step the procedure counted the number of solutions from  $Q$  that fall into each cell by differentiating between: (i) the cells already occupied by population  $P$  (corresponding to vector  $C$  and counted by vector  $Q_c$ ) and (ii) cells only occupied by population  $Q$  (corresponding to vector  $C_q$  and counted by vector  $Q_{cq}$ ). Steps 13-27 in Algorithm 5.

After applying Algorithm 5 to populations  $P$  and  $Q$ , the number of solutions that exists in each cell is stored in vectors  $P_c$ ,  $Q_c$  and  $Q_{cq}$ . To calculate the probability distribution associated with each cell as given by Equation 7.5 one needs to divide the number of solutions counted for each cell by the population size. This last step is implemented directly in Algorithm 6 (defined in Section 7.5) since it simplifies the implementation and speeds up the computation.

## 7.4 Proposed Dissimilarity Measure

In this section, based on the concepts of *entropy* and *relative entropy*, and the multidimensional histogram algorithm introduced earlier, a *dissimilarity* measure is proposed. By comparing an MOEA's populations in successive generations, this measure aims to identify whether or not an MOEA population has stabilised—that is, ceased to significantly vary. If and when an MOEA population stabilises, the *dissimilarity* measure tends to a constant value—zero in the case when the population converges to the POF; and a value greater than zero if the population stagnates away from the POF.

The proposed *dissimilarity* measure is an aggregation of the *dissimilarity* measures corresponding to the intersection non-intersection sets, as discussed below:

1. For the intersection set ( $J_I$ ): accounting for the fact that for a cell  $x_i \in J_I$ , solutions from both  $P$  and  $Q$  exist, the *dissimilarity* measure corresponding to  $J_I$ , namely  $\mathcal{D}(p, q)_I$ , can be given by

$$\mathcal{D}(p, q)_I = \mathcal{KL}(p||q) + \mathcal{KL}(q||p) \quad (7.8)$$

where

$$\mathcal{KL}(p||q) = - \sum_{x_i \in J_I} \frac{p(x_i)}{2} \log \left\{ \frac{q(x_i)}{p(x_i)} \right\} \quad (7.9)$$

and

$$\mathcal{KL}(q||p) = - \sum_{x_i \in J_I} \frac{q(x_i)}{2} \log \left\{ \frac{p(x_i)}{q(x_i)} \right\} \quad (7.10)$$

In this measure the order of the operands, i.e.  $p$  and  $q$ , does not influence the final result since the summation of  $\mathcal{KL}(p||q)$  and  $\mathcal{KL}(q||p)$  makes it a commutative measure, i.e.,  $\mathcal{D}(p, q)_I = \mathcal{D}(q, p)_I$ .

2. For the non-intersection sets ( $P_{NI}$  and  $Q_{NI}$ ): the *dissimilarity* measure, namely  $\mathcal{D}(p, q)_Y$ ,

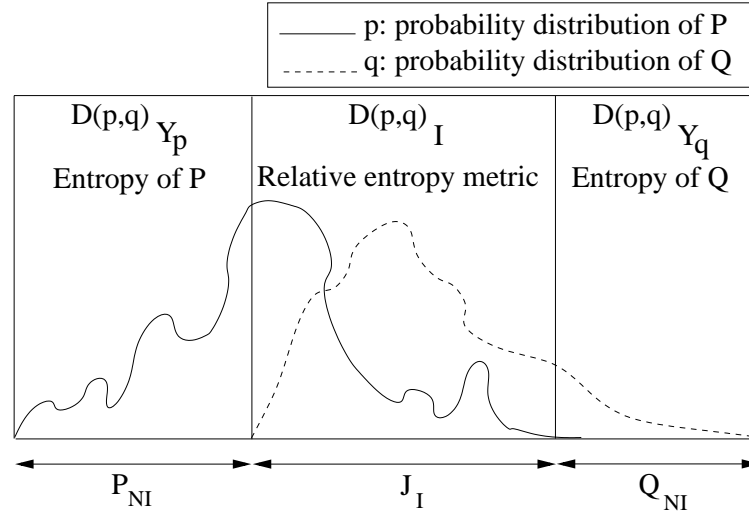


Figure 7.3: Representation of intersection and non-intersection regions obtained for populations  $P$  and  $Q$  accordingly to their probability distribution functions.

can be given by

$$\mathcal{D}(p, q)_Y = \mathcal{D}(p, q)_{Y_P} + \mathcal{D}(p, q)_{Y_Q} \quad (7.11)$$

where

$$\mathcal{D}(p, q)_{Y_P} = - \sum_{x_i \in P_{NI}} \frac{p(x_i)}{2} \log p(x_i) \quad (7.12)$$

and

$$\mathcal{D}(p, q)_{Y_Q} = - \sum_{x_i \in Q_{NI}} \frac{q(x_i)}{2} \log q(x_i) \quad (7.13)$$

Notably, the contributions  $\mathcal{D}(p, q)_{Y_P}$  and  $\mathcal{D}(p, q)_{Y_Q}$ , correspond to the instances where a cell  $x_i \in P_{NI}$  and  $x_i \in Q_{NI}$ , respectively.

In the wake of the above, the proposed *dissimilarity* measure between two MOEA populations, denoted by  $\mathcal{D}(p, q)$ , schematically presented in Figure 7.3, is given by

$$\mathcal{D}(p, q) = \mathcal{D}(p, q)_I + \mathcal{D}(p, q)_Y. \quad (7.14)$$

Notably, the proposed *dissimilarity* measure has the following characteristics: (i) it is non-negative, since  $\mathcal{D}(p, q) \geq 0$ , (ii) it is symmetric, since  $\mathcal{D}(p, q) = \mathcal{D}(q, p)$ , and (iii) it does

not violate the coincidence axiom, since  $\mathcal{D}(p, q) = 0$  iff  $p(X) = q(X), \forall X$ .

Substitution of Equations 7.8 and 7.11 in Equation 7.14 reveals that:

1. if  $p(X) = q(X) \forall X$ , that is: (i)  $p(x_i) = q(x_i) \forall x_i \in J_I$ , (ii)  $P_{NI} = \emptyset$ , and (iii)  $Q_{NI} = \emptyset$ , then  $\mathcal{D}(p, q) = 0$ . In one equation this can be expressed as:

$$p(X) = q(X) \forall X \implies \mathcal{D}(p, q) = 0. \quad (7.15)$$

In other words, when  $P$  and  $Q$  are probabilistically equal within the intersection region, and the non-intersection regions are empty, the proposed *dissimilarity* measure achieves a zero value.

2. if  $p(X) \neq q(X) \forall X$ , then nothing can be said about the value of  $\mathcal{D}(p, q)$ , just that its magnitude will grow as more and more data points fall into the non-intersection region.

In the current context, where the data sets, namely  $P$  and  $Q$ , represent MOEA populations in successive generations, the following inferences can be made:

- (i) during the initial generations, when an MOEA's populations are significantly dissimilar, a significantly high value of *dissimilarity* measure may be obtained.
- (ii) the *dissimilarity* measure may achieve a zero (or approximately zero) value when an MOEA's population over successive generations becomes identical. This may happen when an MOEA converges to the POF (in entirety or in part), or when an MOEA completely stagnates (in the same part of the objective space) away from the POF.
- (iii) the *dissimilarity* measure may continue to retain a constant high value over successive generations, if for some reason the corresponding populations remain identically dissimilar.



Given the state-of-the-art, where most existing MOEAs are known to: (i) provide a good POF-approximation in the case of two- and three-objective problems, the proposed *dissimilarity* measure may achieve a value close to zero, and (ii) fail in providing a good POF-approximation in the case of MaOPs, the proposed *dissimilarity* measure may retain a constant high value. In either case, the underlying MOEA could be terminated because the quality of POF-approximation can not be further improved, despite additional computational expense.

## 7.5 MOEA-termination Algorithm

This section proposes an algorithm for the termination of an MOEA, and refers to it as an MOEA-termination algorithm (Algorithm 6). This algorithm is based on integrating:

- (a) the probability density estimation by a multihistogram algorithm (Algorithm 5),
- (b) the computation of the proposed *dissimilarity* measure (Equation 7.14), and
- (c) a termination criterion based on the mean and standard deviation of the *dissimilarity* measures—defined upto a pre-specified number of decimal places ( $n_p$ ), for a pre-specified number of successive generations ( $n_s$ ) of the MOEA.

While the Items [a] and [b] have previously been discussed in detail, the rationale for the termination criterion mentioned in Item [c] is as follows. MOEAs are stochastic in nature, and hence, even when an MOEA has converged to the POF, some variations in the *dissimilarity* measures may be observed in successive generations (depending on the nature of the underlying problem, and the parameter settings of the variation operators of the underlying MOEA). Towards negating the effect of such variations, and proposing a robust criterion for termination, the use of mean (Equation 7.16) and standard deviation (Equation 7.17) of the *dissimilarity* measures from the initial ( $i = 1$ ,  $i$  denoting the generation counter) to the current generation ( $i = t$ ) is proposed. In that, when the mean and standard

deviation of the *dissimilarity* measures in a pre-specified number of successive generations ( $n_s$ ) of the MOEA may coincide upto a pre-specified number of decimal places ( $n_p$ ), then it is proposed that the underlying MOEA be terminated, and the last generation be reported as  $N_{gt}$ . Notably, the degree of accuracy could be controlled by the user through the choice of  $n_p$  and  $n_s$ .

$$M_t = \frac{1}{t} \sum_{i=1}^t \mathcal{D}_i = \frac{1}{t} \{(t-1) \times M_{t-1} + \mathcal{D}_t\} \quad (7.16)$$

$$S_t = \frac{1}{t} \sum_{i=1}^t (\mathcal{D}_i - M_t)^2 = \frac{1}{t} \{(t-1) \times S_{t-1} + \mathcal{D}_t^2 - M_t^2 + M_{t-1}^2\} \quad (7.17)$$

For clarity, the steps of the proposed Algorithm 6, are as summarised below.

1. Generate a population of feasible non-dominated solutions randomly and let this population be denoted by  $P$ .
2. Run an MOEA for one generation using  $P$  as the initial population, and obtain a new population of feasible non-dominated solutions, say  $Q$ .
3. Using  $n_{bin}$  and populations  $P$  and  $Q$  as input, run Algorithm 5 and obtain the following vectors:
  - (a) a vector that stores the cells corresponding to regions  $J_I$  and  $P_{NI}$ , denoted by  $C$ .  
For each cell in  $C$  let  $P_c$  and  $Q_c$  store the number of solutions corresponding to populations  $P$  and  $Q$ , respectively;
  - (b) a vector that stores the cells corresponding to region  $Q_{NI}$ , denoted by  $C_q$ . For each cell in  $C_q$  let  $Q_{cq}$  store the number of solutions corresponding to population  $Q$ .
4. Initialise the *dissimilarity* measure after the first generation with a zero value, i.e.,  $D_t = 0$  at  $t = 1$ .
5. For each cell in  $C$  proceed as follows:

**Algorithm 6:** MOEA-termination Algorithm

---

**Input:**  
 $n_s$ : number of successive generations of an MOEA for which the mean and standard deviation of the *dissimilarity* measures are to be compared  
 $n_p$ : number of decimal places to which the mean and standard deviation of the *dissimilarity* measures are to be compared  
 $n_{bin}$ : number of bins for the multidimensional histogram  
 $t = 1, M_0 = 0, S_0 = 0, c_1 = false$  and  $c_2 = false$ .

```

1 begin
2   Generate a population of feasible non-dominated solutions randomly and let this population be
   denoted by  $P$ .
3   Run an MOEA for one generation, using  $P$  as initial, and generate a new feasible non-dominated
   population. Let this new population be denoted by  $Q$ .
4    $(C, C_q, P_c, Q_c, Q_{cq}) \leftarrow MultiHistogram(P, Q, n_{bin})$  /* (Algorithm 5) */
5    $\mathcal{D}_t = 0$  /* Initialise the dissimilarity measure at instant  $t$  */
6   for each  $i$  in  $C$  do
7      $p = P_{c,i}/|P|$  /*  $p(x_i), x_i \in J_I$  */
8      $q = Q_{c,i}/|Q|$  /*  $q(x_i), x_i \in J_I$  */
9     if  $q > 0$  then /* Relative entropy */
10       $\mathcal{D}_t = \mathcal{D}_t - \left[ \left( \frac{p}{2} \log \frac{p}{q} \right) + \left( \frac{q}{2} \log \frac{p}{q} \right) \right]$  /* (Equation 7.8) */
11    else if  $q = 0$  then /* Entropy of  $P$  */
12       $\mathcal{D}_t = \mathcal{D}_t - p \log p$  /* (Equation 7.12) */
13    end
14  end
15  for each  $i$  in  $C_q$  do /* Entropy of  $Q$  */
16     $q_{cq,i} = Q_{cq,i}/|Q|$  /*  $q(x_i), x_i \in Q_{NI}$  */
17     $\mathcal{D}_t = \mathcal{D}_t - q \log q$  /* (Equation 7.13) */
18  end
19   $M_t = \frac{M_{t-1} \times (t-1) + \mathcal{D}_t}{t}$  /* Equation 7.16 */
20   $S_t = \frac{t-1}{t} S_{t-1} + \frac{\mathcal{D}_t^2 - M_t^2 + M_{t-1}^2}{t}$  /* Equation 7.17 */
21   $\hat{M}_t = Round(M_t, n_p)$  /* Round  $M_t$  to the  $n_p^{th}$  decimal place */
22   $\hat{S}_t = Round(S_t, n_p)$  /* Round  $S_t$  to the  $n_p^{th}$  decimal place */
23  if  $t > n_s$  then
24    if  $[\hat{M}_t = \hat{M}_{t-1} = \dots = \hat{M}_{t-n_s}]$  then  $c_1 = true$ 
25    if  $[\hat{S}_t = \hat{S}_{t-1} = \dots = \hat{S}_{t-n_s}]$  then  $c_2 = true$ 
26  end
27  if  $c_1 = true$  and  $c_2 = true$  then
28    Report  $Q$  as the final population and set  $N_{gt} = t$ 
29    Terminate the run.
30  else
31    Set  $t = t + 1, c_1 = false, c_2 = false, P = Q$  and go to step 3.
32  end
33 end

```

---

- (a) compute the probability distribution (given by Equation 7.5) associated with populations  $P$  and  $Q$  by using the number of solutions stored in vectors  $P_c$  and  $Q_c$ , respectively;
  - (b) using the probability distributions obtained above, compute the *dissimilarity* measure by: (i) using Equation 7.8 if the cell belongs to the intersection region, i.e.,  $J_I$  or (ii) using Equation 7.12 if the cell belongs to the non-intersection region, i.e.,  $P_{NI}$ . For both cases add the obtained value to  $\mathcal{D}_t$ .
6. For each cell in  $C_q$  proceed as follows:
- (a) compute the probability distribution (given by Equation 7.5) associated with population  $Q$  by using the number of solutions stored in vector  $Q_{cq}$ ;
  - (b) using the probability distribution obtained above, compute the *dissimilarity* measure by using Equation 7.13 and add the obtained value to  $\mathcal{D}_t$ .
7. Compute the mean ( $M_t$ ) and the standard deviation ( $S_t$ ) of the *dissimilarity* measure at instant  $t$ , by using Equations 7.16 and 7.17, respectively. Subsequently, round-off both  $M_t$  and  $S_t$  to the  $n_p^{th}$  decimal place and store the values into  $\hat{M}_t$  and  $\hat{S}_t$ .
8. When the number of generations ( $t$ ) exceed  $n_s$ , compare the values stored by  $\hat{M}_t$  and  $\hat{S}_t$  between the generations  $t$  and  $t - n_s$ , and proceed as follows:
- (a) if the values between  $\hat{M}_t$  and  $\hat{M}_{t-n_s}$  are equal then set  $c_1 = true$ ;
  - (b) if the values between  $\hat{S}_t$  and  $\hat{S}_{t-n_s}$  are equal then  $c_2 = true$ .

Finally, if both  $c_1 = true$  and  $c_2 = true$ , terminate the MOEA and declare  $N_{gt} = t$ ; else, set  $P = Q$  and return to Step 2.

It may be noted that for a particular choice of the parameter  $n_{bin}$ , the value of  $N_{gt}$  as determined by the proposed algorithm will vary with the other parameters involved, namely,  $n_s$  and  $n_p$ . In that, the value of  $N_{gt}$  will be higher corresponding to higher values of  $n_s$

and  $n_p$ , implying a higher computational cost but a more reliable inference on whether a good POF-approximation has been obtained by the underlying MOEA, or that it can not be obtained (for a given problem, and given parameter settings for the chosen MOEA).

## 7.6 Test Problems and Experimental Settings

The effectiveness of the proposed MOEA-termination algorithm is tested against a wide range of standard test problems (both redundant and non-redundant), involving a varying number of objectives (two to 50), constraints, and variables. The details of these problems, and the obtained results are presented in this section.

### 7.6.1 Test problems

The details of the test problems considered, are summarised in Table 7.1.

### 7.6.2 Experimental Settings

To test the effectiveness of the proposed MOEA-termination algorithm (Algorithm 6), NSGA-II has been chosen as the underlying MOEA, with the following parameter settings. With a population size of 200, the probability of crossover and mutation used is 0.9 and 0.1, respectively, while the distribution index for crossover and mutation is chosen as 5 and 20, respectively.

For each problem, the simulations are performed corresponding to 10 different solution sets obtained from NSGA-II, in that: (i) the number of bins (in multidimensional histogram algorithm) is chosen as 10, i.e.,  $n_{bin} = 10$ , (ii) the mean and standard deviation of the *dissimilarity* measures are compared for 20 successive generations of NSGA-II, i.e.,  $n_s = 20$ , and (iii) the results are reported for three different precision levels, given by  $n_p = 2, 3$  and 4.

The rationale for the choice of the parameters relating to the MOEA-termination algorithm, namely,  $n_s$ ,  $n_p$ , and  $n_{bin}$  lies in the following. MOEAs are stochastic in nature

Table 7.1: Test-suite for the MOEA-termination algorithm (Algorithm 6).

Category	Problem(s)	Objectives ( $M$ )	Number of Constraints ( $J$ )	Variables ( $n$ )
Non-redundant	<u>CTP</u> [28]			
	CTP1	2	2	2
	CTP2 to CTP7	2	1	2
	<u>ZDT</u> [26]			
	ZDT1 to ZDT3	2	0	30
	ZDT4	2	0	10
	ZDT5	2	0	11
	ZDT6	2	0	10
	BNH [172]	2	2	2
	FON [147]	2	0	5
	KUR [173]	2	0	3
	OSY [150]	2	6	6
	POL [174]	2	0	2
	SCH1 & SCH2 [175]	2	0	1
	SRN [120]	2	2	2
	TNK [149]	2	2	2
	VNT [176]	3	0	2
	<u>DTLZ</u> [27]			
	DTLZ1	5,15,25	0	$M + 4$
	DTLZ2 to DTLZ4	5,15,25	0	$M + 9$
Redundant	<u>DTLZ5(<math>I, M</math>)</u> [177]			
	$I = 2$	5,20,50		
	$I = 3$	5,20		
	$I = 5$	10,20	$M - I + 1$	$M + 9$
	$I = 7$	10,20		

where they evolve a randomly initialised population through the *variation* operators (such as crossover and mutation) and *selection* of *fitter* population members, over the generations. Hence, unless an MOEA has converged to the POF or it has stagnated in terms of its ability to further improve the POF-approximation, a significant variation in the population obtained in successive generations is expected. In such a case, the mean and standard deviation of the *dissimilarity* measures in successive generations may not closely conform with each other. In contrast, if the mean and standard deviation of the *dissimilarity* measures computed up to two, three or four decimal places ( $n_p$ ) may conform with each other over some successive generations ( $n_s$ ), then this could be treated as a good indicator that the MOEA population has either converged to the POF or has stagnated such that no further

improvement in the quality of POF-approximation may be possible. Clearly, corresponding to  $n_p = 4$ , a smaller  $n_s$  may be sufficient, than that required corresponding to  $n_p = 2$ . However, for simplicity, this thesis assumes that  $n_s = 20$  (a reasonably high number) may be a reliable choice for  $n_p = 2, 3$  and 4.

With regard to  $n_{bin}$  that controls the number of cells into which the search space will be partitioned, its optimal value should be estimated for a given problem and an MOEA. Notably, in the general context of probability density estimation, efforts have been made to determine a suitable value of  $n_{bin}$  [178, 179, 180, 181]. However, in this thesis such approaches have not been utilised owing to the following reasons:

- estimation of an optimal value of  $n_{bin}$  is in itself a computationally expensive task, and in the context of multi-objective optimisation, where either or both the number of objectives and the population size may be large, the estimation of  $n_{bin}$  may not be computationally feasible.
- in many applications, the data points may be scattered all over the search space. However, in the context of multi-objective optimisation, these data points are non-dominated solutions which in principle capture the trade-off between the objectives involved. In other words, these non-dominated solutions are expected to occupy only a small fraction of the total number of cells determined by  $n_{bin}$ . Furthermore, it needs to be recognised that an optimal value of  $n_{bin}$  with regard to the probability density estimation may not necessarily be optimal with regard to the estimation of the difference in the density functions of two different data sets, as is the case with the proposed *dissimilarity* measure.

Besides the above, it needs to be recognised that: (i) the termination of the underlying MOEA by the proposed Algorithm 6 relies purely on assessing the variation in the mean and standard deviation of the *dissimilarity* measure, and (ii) only the interpretation (post termination) of whether the MOEA has converged or stagnated is based on the absolute

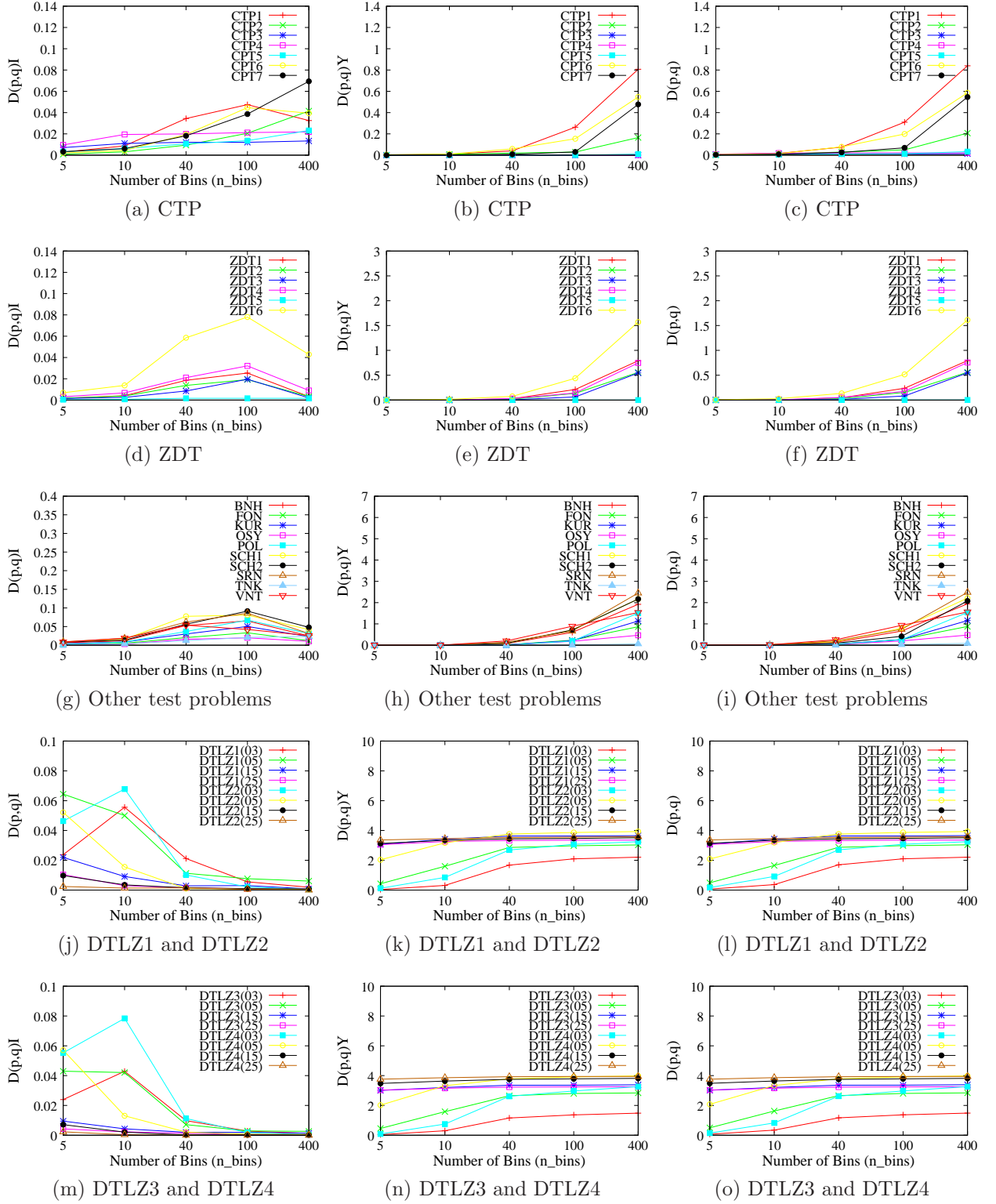


Figure 7.4: Non-redundant test problems: Influence of  $n_{bin}$  on the proposed *dissimilarity* measure ( $\mathcal{D}(p, q)$ ) comprising of  $\mathcal{D}(p, q)_I$  (for the intersection region) and  $\mathcal{D}(p, q)_Y$  (for the non-intersection region). Each plot in this figure is averaged over 10 different solution sets obtained from NSGA-II runs, corresponding to a population size of 200 and 1000 generations.



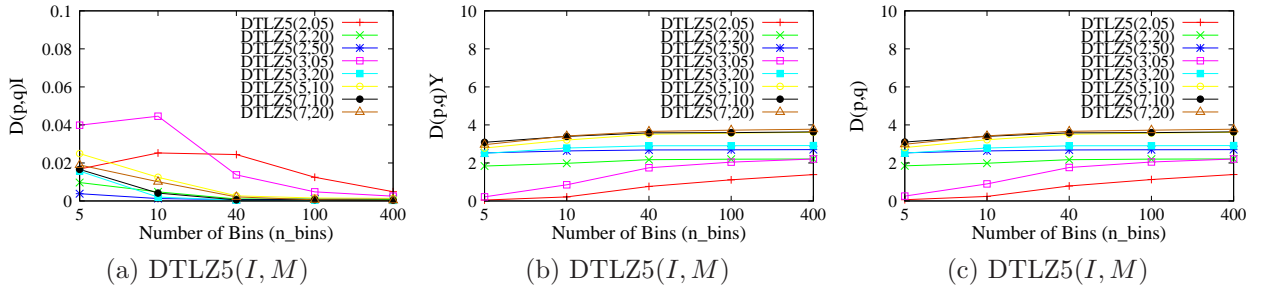


Figure 7.5: Redundant test problems: Influence of  $n_{bin}$  on the proposed *dissimilarity* measure ( $\mathcal{D}(p, q)$ ) comprising of  $\mathcal{D}(p, q)_I$  (for the intersection region) and  $\mathcal{D}(p, q)_Y$  (for the non-intersection region). Each plot in this figure is averaged over 10 different solution sets obtained from NSGA-II runs, corresponding to a population size of 200 and 1000 generations.

value of the *dissimilarity* measure. Notably, if  $n_{bin}$  is fixed to a value that is:

- a too small: then the solutions from populations  $P$  and  $Q$  which may be significantly different may end up sharing a common cell. This anonymity is more likely to manifest itself in many-objective problems ( $M \gg 4$ ), where the search space may need to be partitioned in to a higher number of cells. In such problems, a falsely deflated value of the *dissimilarity* measure may be obtained, as the intersecting and the non-intersecting region may erroneously become larger ( $\mathcal{D}(p, q)_I$ ) and smaller ( $\mathcal{D}(p, q)_Y$ ), respectively.
- b too high: it is likely that two solutions—one each from populations  $P$  and  $Q$ , may fail to share the same cell despite being significantly similar (or even near identical). This anonymity is more likely to manifest itself in two- and three-objective problems, where partitioning the search space in a relatively smaller number of cells may be sufficient. In such problems, a falsely inflated value of the *dissimilarity* measure may be obtained, as the intersecting and the non-intersecting region may erroneously become smaller  $\mathcal{D}(p, q)_I$  and larger  $\mathcal{D}(p, q)_Y$ , respectively.

Experiments are performed on the test-suite highlighted above, with  $n_{bin}$  varying from 2 to 400, and the effect of this variation on  $\mathcal{D}(p, q)_I$ ,  $\mathcal{D}(p, q)_Y$ , and  $\mathcal{D}(p, q)$  is presented in Figure 7.4 and 7.5. Clearly, these results for a wide range of test problems (in terms of the number of objectives, constraints, variables, and redundant/non-redundant characteristics):

(i) support the arguments presented in Items [a] and [b] above, and (ii) reveal that  $n_{bin} = 10$  should be a good choice for the entire test-suite in consideration.

## 7.7 Experimental Results

This section reports the performance of the proposed MOEA-termination algorithm (Algorithm 6) on the test-suite presented above, in the following forms:

- $N_{gt}$  determined by Algorithm 6 is reported for  $n_p = 2, 3, 4$  ( $n_{bin} = 10$  and  $n_s = 20$ ) in a tabular form.
- the populations obtained from NSGA-II at  $N_{gt}$  corresponding to  $n_p = 2, 3, 4$  are plotted together to facilitate a comparison between them.
- the absolute value of the *dissimilarity* measure ( $\mathcal{D}(p, q)$ ) is plotted along with the corresponding mean and standard deviation, over as many generations as deemed necessary to plot  $N_{gt}$  corresponding to  $n_p = 4$  (requiring more generations compared to  $n_p = 2, 3$ ).
- to facilitate a better visualisation of how  $\mathcal{D}(p, q)$  varies from the initial generation onwards, its snapshot for the first 100 generations is presented.

### 7.7.1 Experimental Results for Two- and Three-objective Problems

The  $N_{gt}$  determined by Algorithm 6 is reported in Table 7.2, where the broader trend is that the  $\mathcal{D}(p, q)$  values for all the problems considered, are reasonably close to zero, regardless of the value of  $n_p$ . In the light of Equation 7.15, this trend indicates that NSGA-II may have converged for these problems. This argument is largely validated through the corresponding population plots in Figure 7.6 to Figure 7.12. Notably, as the value of  $n_p$  is raised from 2 to

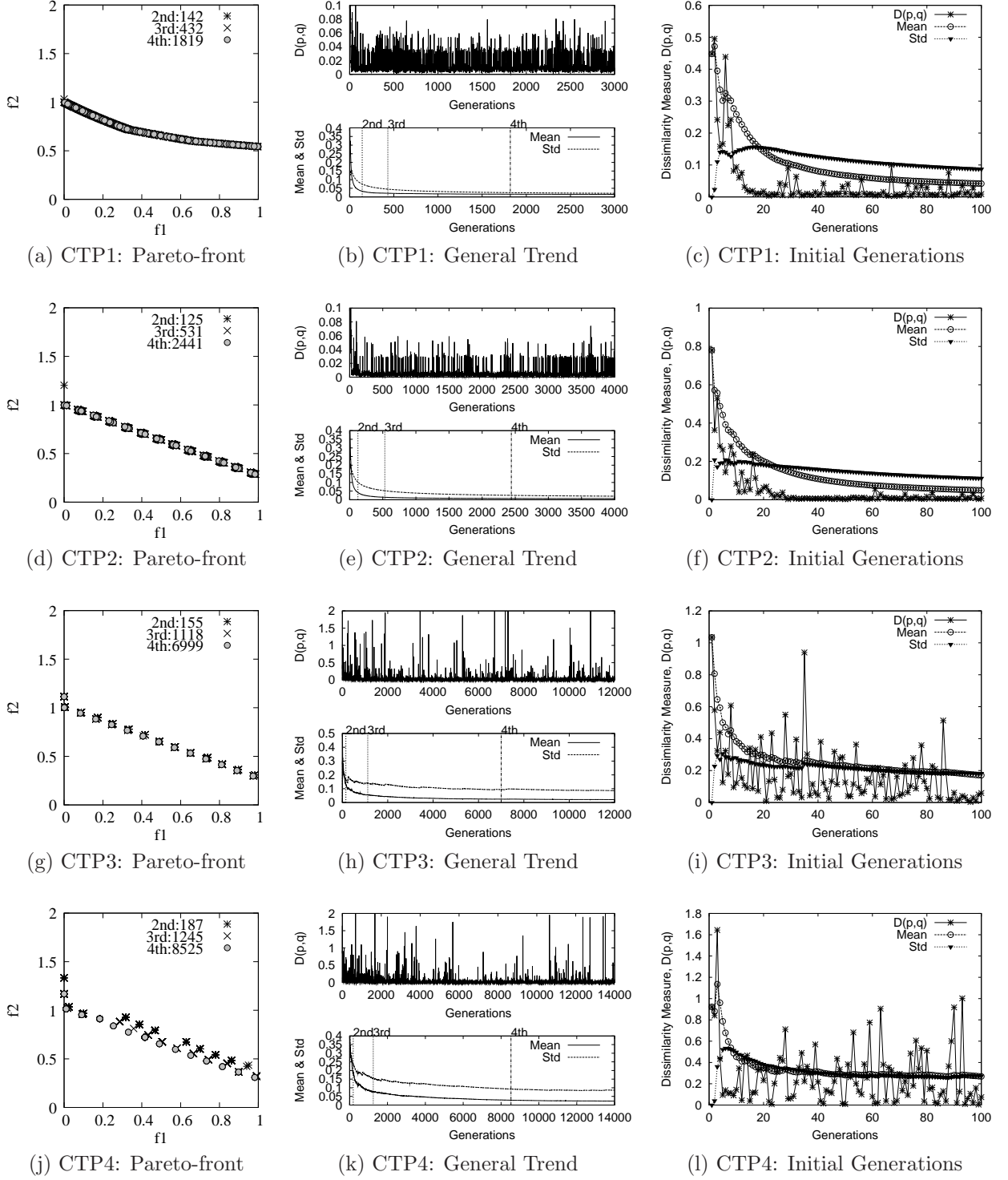


Figure 7.6: CTP1 to CTP4: Performance evaluation of Algorithm 6 through the NSGA-II populations at  $N_{gt}$  corresponding to  $n_p = 2, 3, 4$ ; and  $\mathcal{D}(p, q)$  along with its mean and standard deviation, and a zoomed snapshot limited to the first 100 generations.

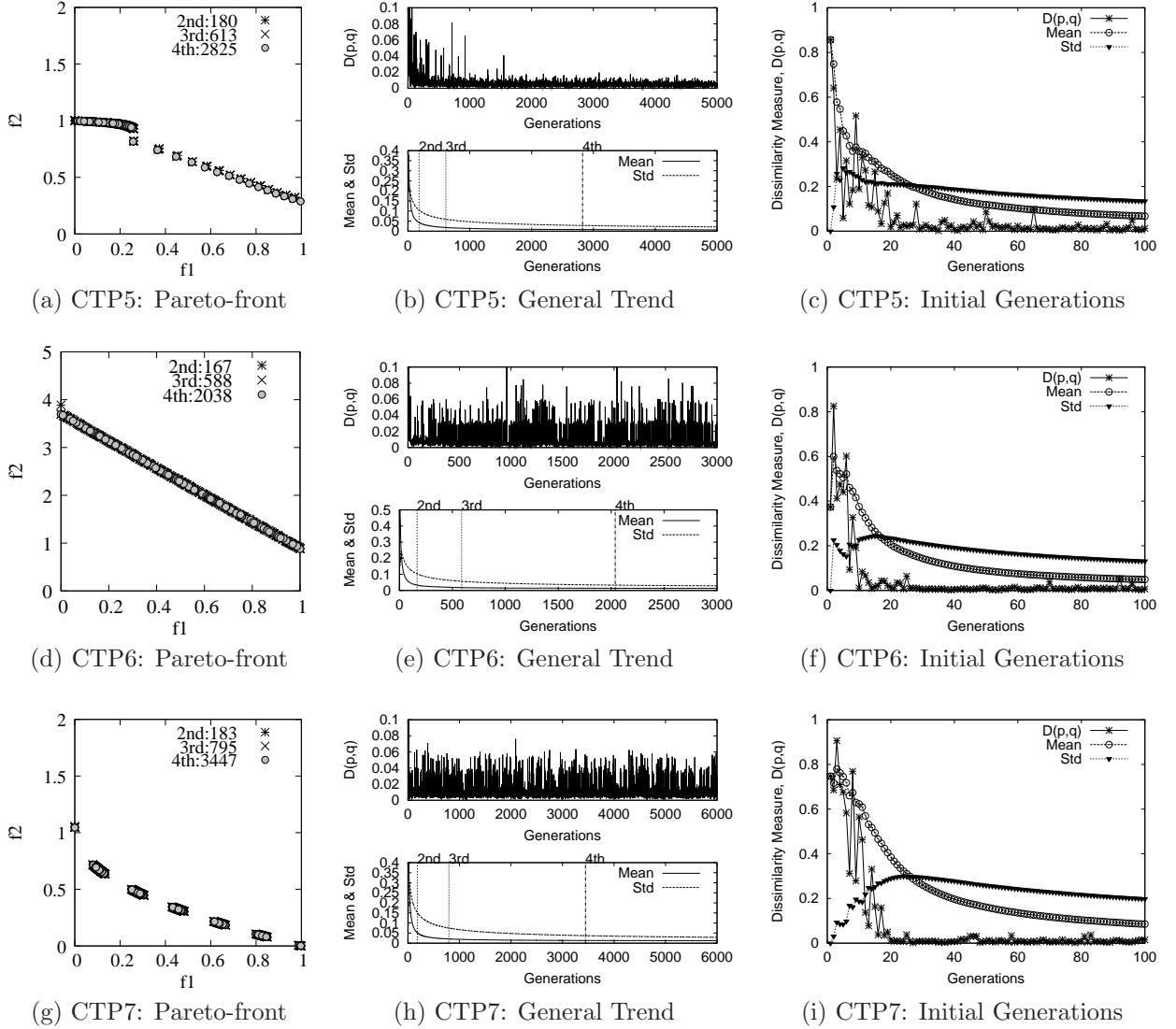


Figure 7.7: CTP5 to CTP7: Performance evaluation of Algorithm 6 through the NSGA-II populations at  $N_{gt}$  corresponding to  $n_p = 2, 3, 4$ ; and,  $\mathcal{D}(p, q)$  along with its mean and standard deviation, and a zoomed snapshot limited to the first 100 generations.

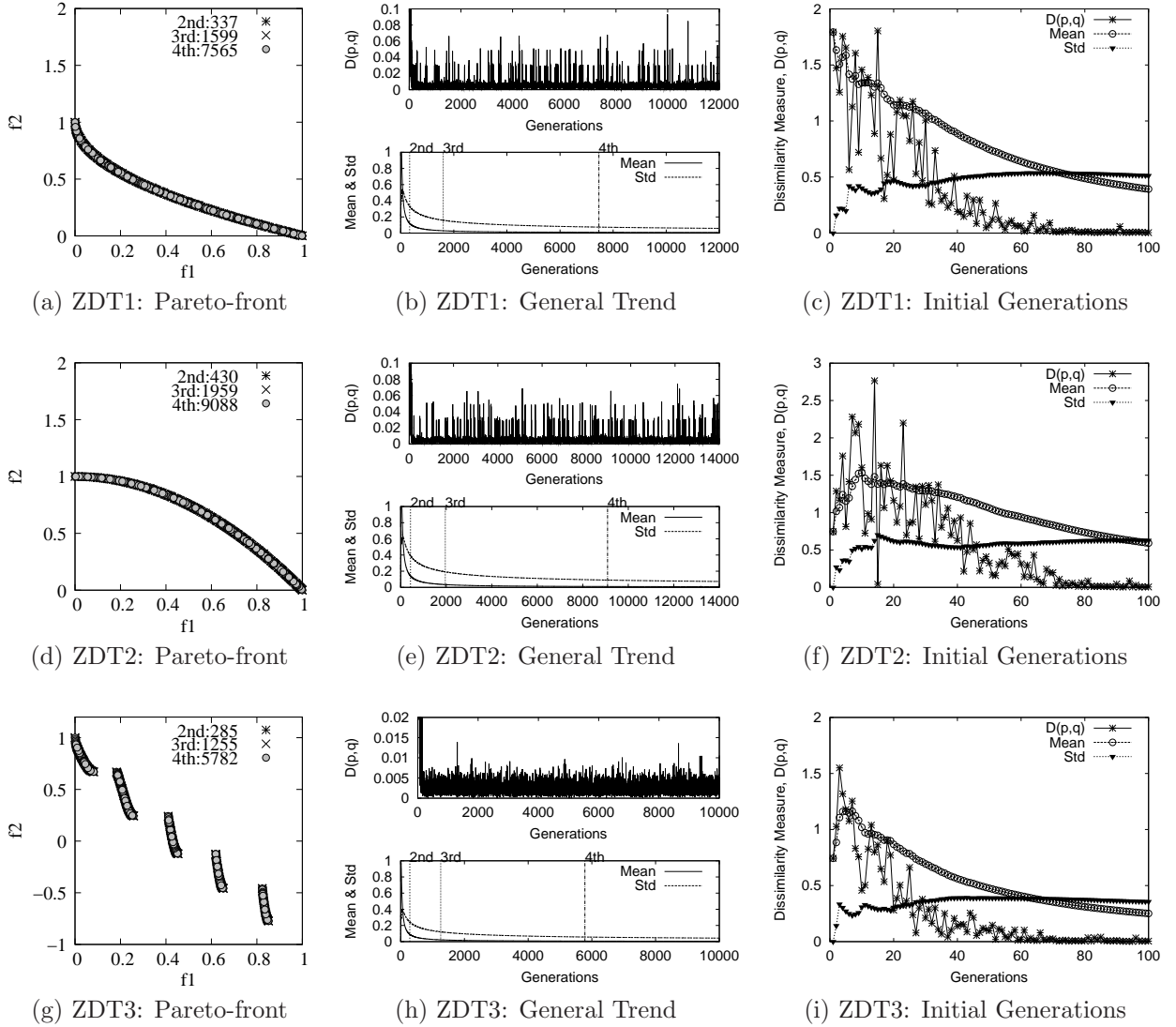


Figure 7.8: ZDT1 to ZDT3: Performance evaluation of Algorithm 6 through the NSGA-II populations at  $N_{gt}$  corresponding to  $n_p = 2, 3, 4$ ; and  $\mathcal{D}(p, q)$  along with its mean and standard deviation, and a zoomed snapshot limited to the first 100 generations.

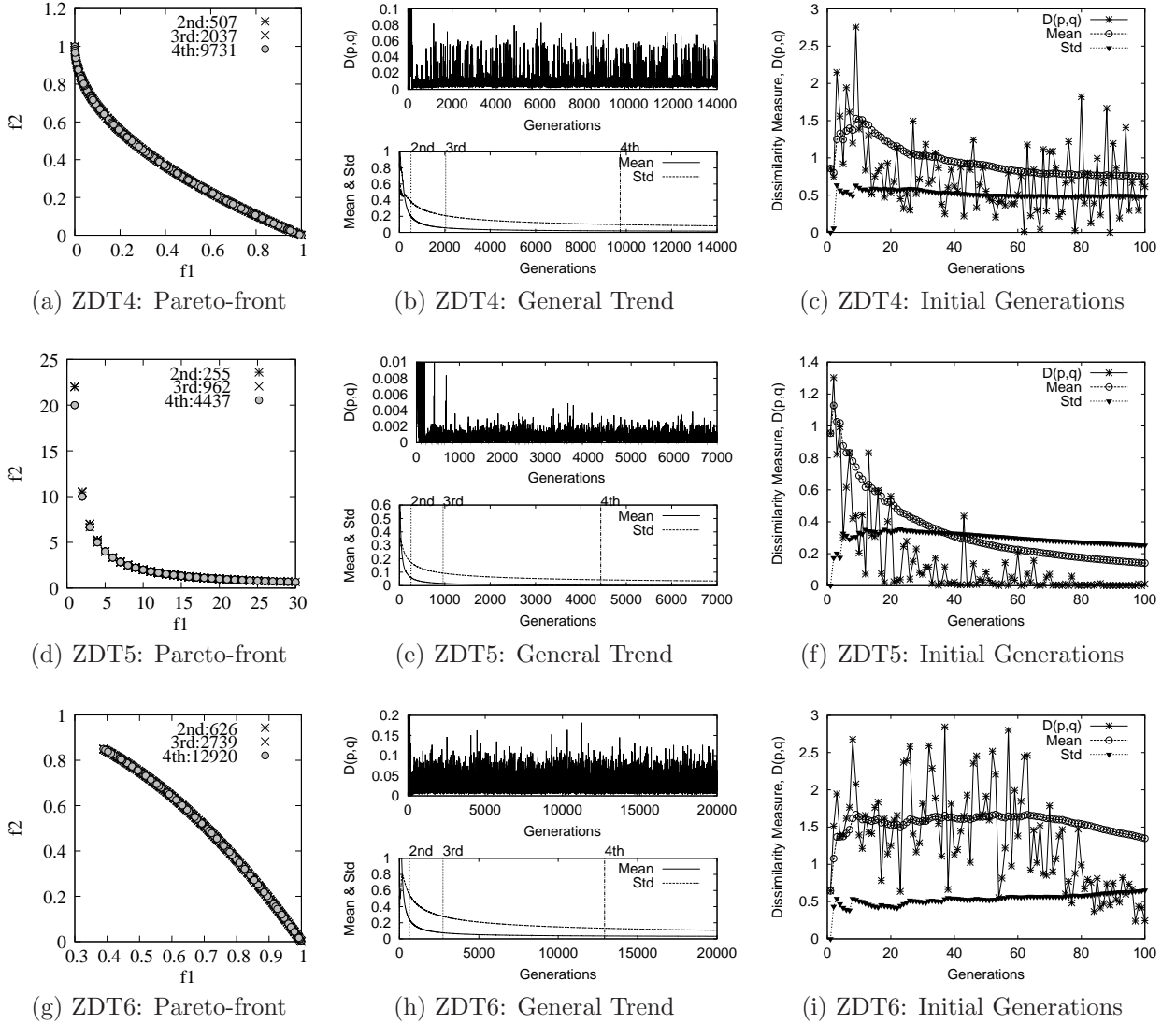


Figure 7.9: ZDT4 to ZDT6: Performance evaluation of Algorithm 6 through the NSGA-II populations at  $N_{gt}$  corresponding to  $n_p = 2, 3, 4$ ; and  $\mathcal{D}(p, q)$  along with its mean and standard deviation, and a zoomed snapshot limited to the first 100 generations.

Table 7.2:  $N_{gt}$  determined by the MOEA-termination algorithm (Algorithm 6) on two- and three-objective test problems, corresponding to  $n_{bin} = 10$ ,  $n_s = 20$  and  $n_p = 2, 3, 4$

Each result is an average over 10 different solution sets obtained from NSGA-II runs, corresponding to a population size of 200. In the case of the CTP and ZDT problem sets, the maximum  $N_{gt}$  reported by a particular problem has been underlined

Problem	Generations $N_{gt}$			Dissimilarity Measure $\mathcal{D}(p, q)$		
	$n_p = 2$	$n_p = 3$	$n_p = 4$	$n_p = 2$	$n_p = 3$	$n_p = 4$
CTP1	150±27	595±148	2399±691	0.035±0.001	0.019±0.001	0.016±0.000
CTP2	133±20	539±71	2316±430	0.036±0.004	0.013±0.001	0.007±0.000
CTP3	176±33	1025±151	6376±830	0.118±0.020	0.049±0.006	0.024±0.001
CTP4	<u>232±38</u>	<u>1314±119</u>	<u>8505±1159</u>	<u>0.144±0.023</u>	<u>0.066±0.010</u>	<u>0.030±0.003</u>
CTP5	160±18	578±40	2672±174	0.044±0.005	0.018±0.001	0.008±0.000
CTP6	149±22	592±122	2259±419	0.033±0.003	0.017±0.001	0.013±0.000
CTP7	196±21	820±74	3625±405	0.045±0.006	0.019±0.001	0.013±0.000
ZDT1	393±26	1771±122	8224±550	0.112±0.007	0.029±0.001	0.010±0.000
ZDT2	462±27	2115±133	9880±582	0.146±0.008	0.035±0.001	0.011±0.000
ZDT3	320±26	1411±104	6480±466	0.092±0.008	0.023±0.001	0.007±0.000
ZDT4	502±26	2167±102	10244±480	0.218±0.010	0.057±0.002	0.019±0.000
ZDT5	246±24	1023±53	4688±284	0.064±0.006	0.016±0.001	0.004±0.000
ZDT6	<u>597±61</u>	<u>2714±41</u>	<u>12658±240</u>	<u>0.248±0.019</u>	<u>0.074±0.001</u>	<u>0.035±0.000</u>
BNH	56±13	148±43	975±185	0.021±0.002	0.020±0.002	0.020±0.000
FON	330±42	1456±178	6664±840	0.059±0.004	0.018±0.001	0.009±0.000
KUR	282±24	1255±102	5666±456	0.052±0.004	0.019±0.001	0.011±0.000
OSY	323±39	1478±157	6772±798	0.084±0.010	0.021±0.002	0.008±0.000
POL	139±19	542±125	2448±581	0.028±0.004	0.014±0.001	0.010±0.000
SCH1	73±12	226±48	1164±128	0.035±0.003	0.031±0.001	0.029±0.000
SCH2	53±19	178±52	860±135	0.024±0.002	0.021±0.001	0.019±0.000
SRN	130±26	530±107	2326±557	0.035±0.003	0.023±0.001	0.020±0.000
TNK	206±37	866±147	3937±695	0.043±0.002	0.013±0.001	0.006±0.000
VNT	156±22	604±98	2415±488	0.054±0.003	0.040±0.002	0.035±0.000

4, the  $\mathcal{D}(p, q)$  values become smaller (tend to zero), though no significant difference could be visually observed in the corresponding population plots, with the exception of the CTP4 problem (discussed further, below).

Given the above observations, it can be inferred that the proposed Algorithm 6 is effective in achieving its goal of identifying  $N_{gt}$  (subject to a specified  $n_p$ ). For instance, in the case of  $n_p = 2$ , it is able to suggest the termination of NSGA-II in as few as  $53 \pm 19$  generations in the case of the SCH2 problem, to as many as  $597 \pm 61$  generations in the case of the ZDT6 problem. The revelation in Figures 7.9g and 7.11g that running NSGA-II any further beyond the identified  $N_{gt}$  does not significantly help improve the quality of POF-approximation

endorses the effectiveness of Algorithm 6. It is outside the scope of this thesis to relate the reported  $N_{gt}$  with the features of the corresponding problems, yet, some discussion in this direction is being made with respect the CTP and ZDT class of problems, below.

It can be seen in Table 7.2 that within the CTP class of problems, the  $N_{gt}$  and  $\mathcal{D}(p, q)$  values corresponding to CTP4 are the highest (regardless of  $n_p$ ). It implies that from the MOEA perspective, NSGA-II in this case, CTP4 happened to be the most difficult problem within the CTP class. This inference could well be justified on the basis of:

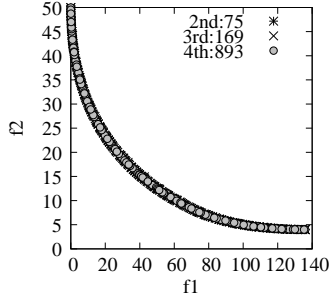
- Problem features: In the case of CTP4, the Pareto-optimal solutions lie at the end of long narrow *tunnels*, which poses difficulties for an MOEA. This also explains why CTP4 is the only problem where significant improvements could be visually observed, as  $n_p$  is raised from 2 to 4 (Figure 7.6j).
- Available literature: In [28], it has been highlighted that the NSGA-II and Ray et al.'s algorithm struggled to approximate the POF.

Similarly, in the case of ZDT problems, it can be seen in Table 7.2 that the  $N_{gt}$  and  $\mathcal{D}(p, q)$  values corresponding to ZDT6 are the highest, followed by ZDT4 (regardless of  $n_p$ ). It implies that within the ZDT class, these problems may be among the more difficult ones. This inference could well be justified based on the fact that compared to other problems:

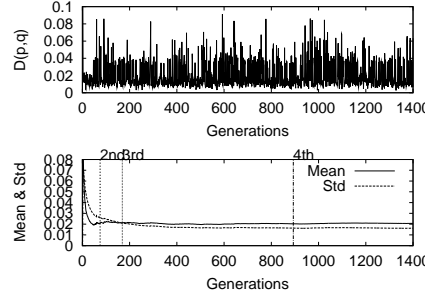
- ZDT6 poses more difficulties to an MOEA owing to the non-uniformity in its search space, in that: (i) the Pareto-optimal solutions are not uniformly distributed across the POF; and (ii) the density of solutions changes with respect to their proximity from the POF, such that the density closer to the POF is lower, which grows with the distance from the POF.
- ZDT4 poses more difficulties to an MOEA owing to the presence of  $21^9$  local Pareto-optimal fronts.

The results for the other (excluding the CTP and ZDT problems) two- and three-objective problems are presented in Table 7.2 and Figures 7.10 to 7.12. The observations that: (i) the

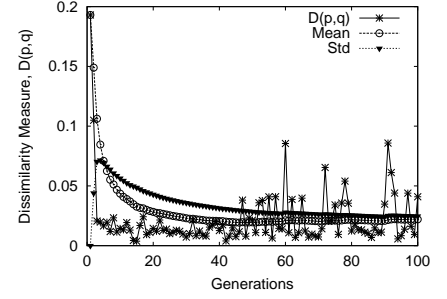




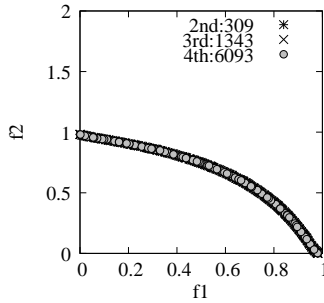
(a) BNH: Pareto-front



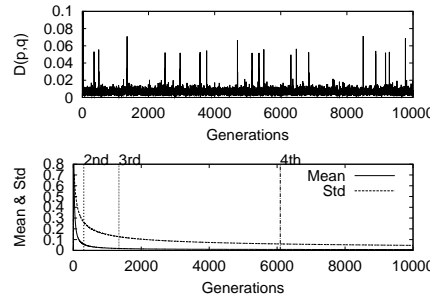
(b) BNH: General Trend



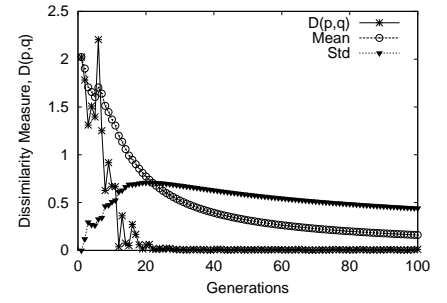
(c) BNH: Initial Generations



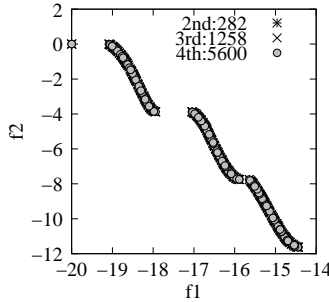
(d) FON: Pareto-front



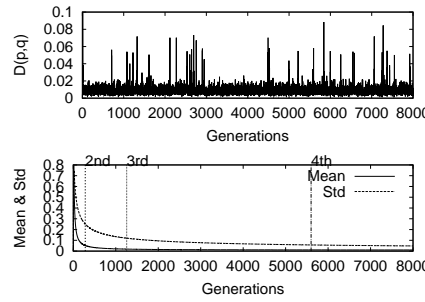
(e) FON: General Trend



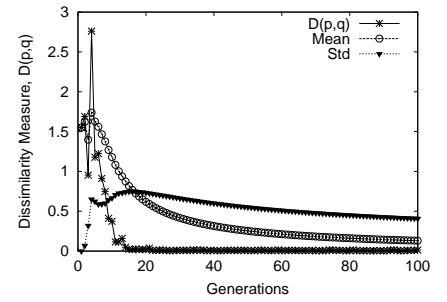
(f) FON: Initial Generations



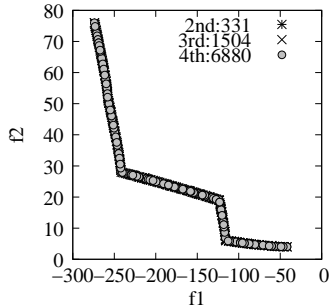
(g) KUR: Pareto-front



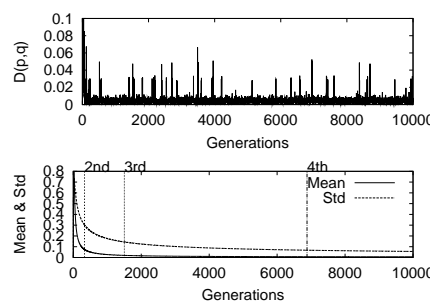
(h) KUR: General Trend



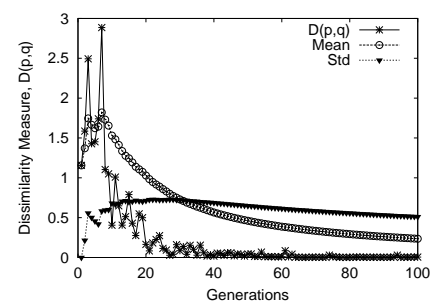
(i) KUR: Initial Generations



(j) OSY: Pareto-front



(k) OSY: General Trend



(l) OSY: Initial Generations

Figure 7.10: BNH, FON, KUR and OSY: Performance evaluation of Algorithm 6 through the NSGA-II populations at  $N_{gt}$  corresponding to  $n_p = 2, 3, 4$ ; and,  $\mathcal{D}(p, q)$  along with its mean and standard deviation, and a zoomed snapshot limited to the first 100 generations.

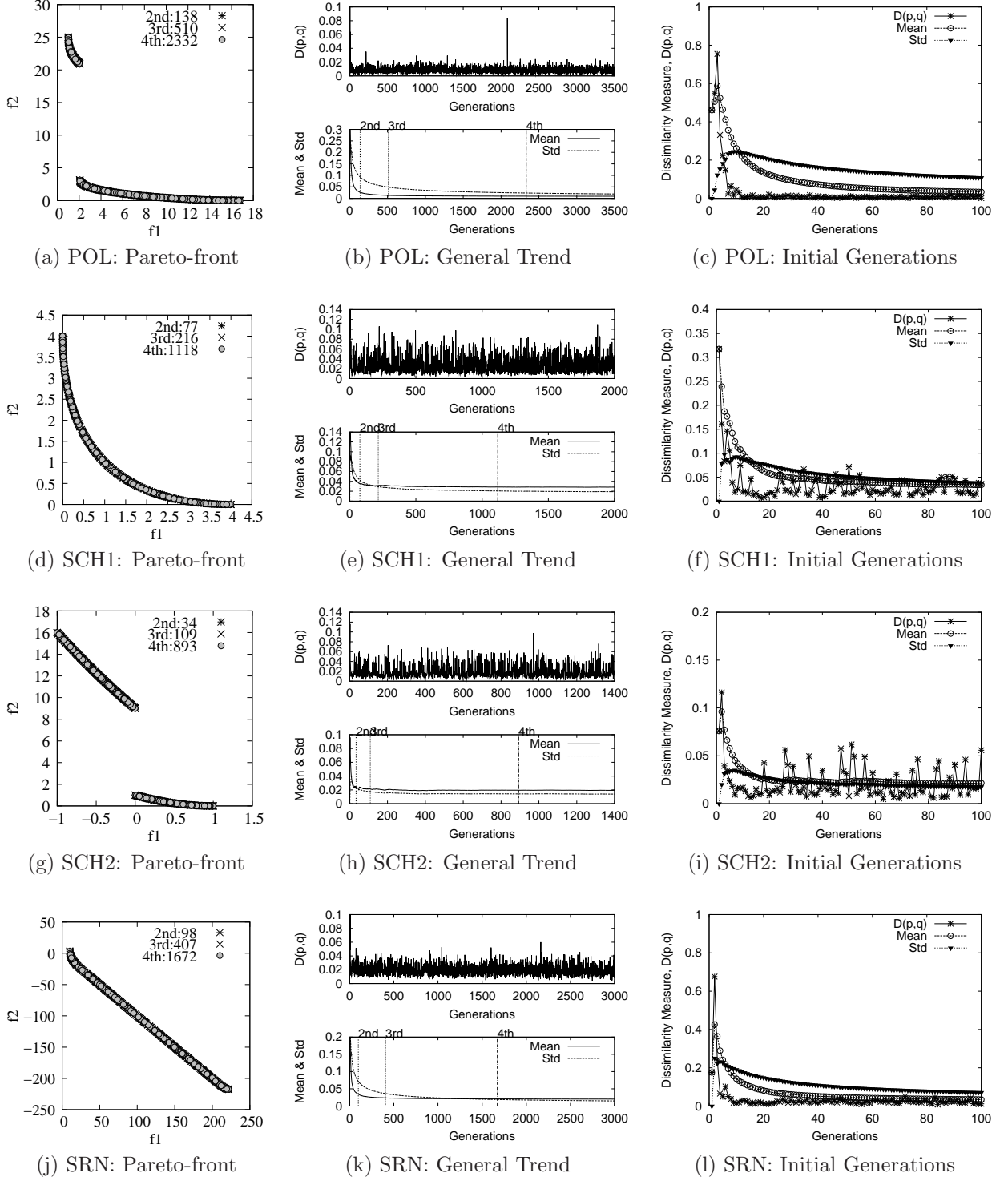


Figure 7.11: POL, SCH1, SCH2 and SRN: Performance evaluation of Algorithm 6 through the NSGA-II populations at  $N_{gt}$  corresponding to  $n_p = 2, 3, 4$ ; and,  $\mathcal{D}(p, q)$  along with its mean and standard deviation, and a zoomed snapshot limited to the first 100 generations.

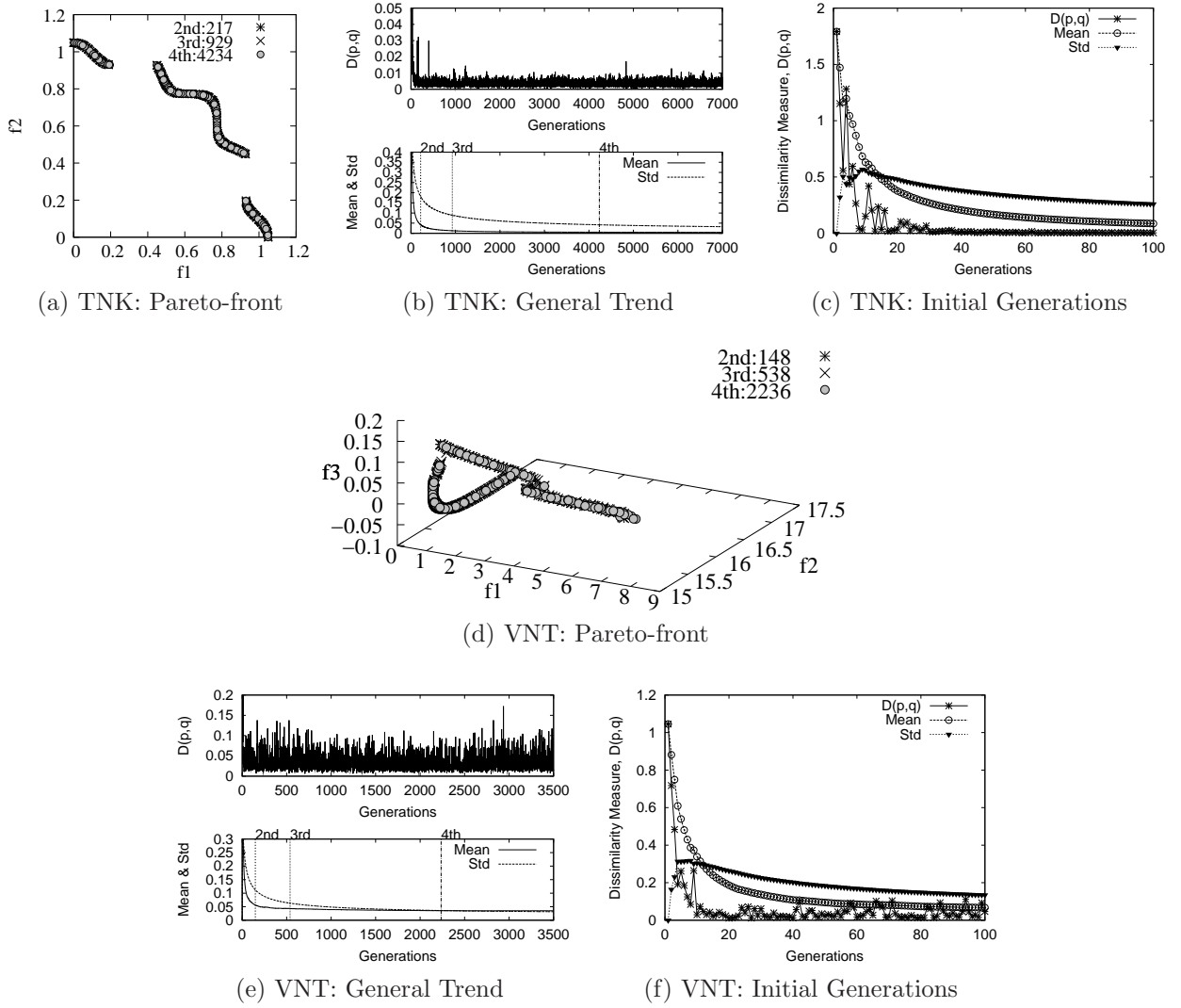


Figure 7.12: TNK and VNT: Performance evaluation of Algorithm 6 through the NSGA-II populations at  $N_{gt}$  corresponding to  $n_p = 2, 3, 4$ ; and,  $\mathcal{D}(p, q)$  along with its mean and standard deviation, and a zoomed snapshot limited to the first 100 generations.

values of  $\mathcal{D}(p, q)$  are reasonably close to zero for each of  $n_p = 2, 3, 4$ , and (ii) the NSGA-II populations at  $N_{gt}$  corresponding to  $n_p = 2, 3, 4$  significantly overlap, are indicative of the fact that NSGA-II is able to converge to the POF for these problems. To add further meaning to these results, the following observations can be made.

- The  $N_{gt}$  corresponding to the SCH1 and SCH2 problems is the lowest (among this set of problems), and this could be attributed to the fact that the nature of objective functions is relatively simpler than other problems. In that, while both the objectives are quadratic in the case of SCH1, SCH2 involves one quadratic and one linear objective function.
- Despite the fact that VNT is a three-objective problem, the  $N_{gt}$  corresponding to it (regardless of  $n_p$ ) is lower than some of the two-objective problems. It is well known that the number of objectives alone do not define the degree of difficulty, it is in fact also guided by the nature of objective functions, and that of the search space. It is interesting to note that the results produced by the proposed Algorithm 6 do not contradict these facts.

### 7.7.2 Experimental Results for Many-objective Test Problems

The results obtained from Algorithm 6 for test MaOPs, namely DTLZ and DTLZ5( $I, M$ ), are presented in Table 7.3, and Figures 7.13 to 7.16. The issue that most existing MOEAs fail to provide a good POF-approximation in the case of MaOPs has been discussed earlier (Chapter 2), in that, most MOEAs stagnate in terms of their ability to improve on the quality of the POF-approximation beyond a certain stage. This observation has been linked (earlier in this chapter) to the possible band of values (close or far from zero) that the measure  $\mathcal{D}(p, q)$  can assume. Given that MOEAs are stochastic in nature, in a scenario where they stagnate: (i) greater variability in the solutions from two successive generations could be observed, compared to the scenario where an MOEA converges to the POF, (ii) solutions

Table 7.3:  $N_{gt}$  determined by the MOEA-termination algorithm (Algorithm 6) for many-objective versions of the DTLZ and DTLZ5( $I, M$ ) problems, corresponding to  $n_{bin} = 10$ ,  $n_s = 20$  and  $n_p = 2, 3, 4$

Each result is an average over 10 different solution sets obtained from NSGA-II runs, corresponding to a population size of 200

Problem	Generations $N_{gt}$			Dissimilarity Measure $\mathcal{D}(p, q)$		
	$n_p = 2$	$n_p = 3$	$n_p = 4$	$n_p = 2$	$n_p = 3$	$n_p = 4$
DTLZ1(05)	357±158	2544±349	17266±4999	1.193±0.242	1.614±0.173	1.617±0.242
DTLZ1(15)	318±75	2236±400	23587±6688	3.031±0.105	3.310±0.121	3.403±0.067
DTLZ1(25)	322±54	1947±507	19358±4530	3.114±0.066	3.248±0.089	3.215±0.096
DTLZ2(05)	216±44	1490±230	10592±1132	3.093±0.038	3.113±0.019	3.161±0.016
DTLZ2(15)	276±33	1557±386	10579±1208	3.517±0.041	3.438±0.052	3.274±0.055
DTLZ2(25)	308±58	1792±278	12912±2224	3.561±0.064	3.401±0.085	3.228±0.036
DTLZ3(05)	336±100	1886±235	11543±1496	1.835±0.086	1.664±0.120	1.633±0.109
DTLZ3(15)	282±40	1893±516	14477±1366	3.298±0.023	3.208±0.108	2.971±0.127
DTLZ3(25)	354±79	2470±933	17671±2946	3.136±0.066	2.992±0.131	2.653±0.213
DTLZ4(05)	420±36	1766±183	10667±1097	3.019±0.032	3.120±0.011	3.135±0.003
DTLZ4(15)	268±40	1299±141	10388±889	3.599±0.020	3.670±0.010	3.621±0.007
DTLZ4(25)	235±65	1422±171	10480±914	3.520±0.043	3.652±0.007	3.652±0.009
DTLZ5(2,05)	464±59	2015±292	8776±1414	0.436±0.043	0.271±0.077	0.225±0.112
DTLZ5(2,20)	706±367	3209±982	15244±1944	2.549±0.476	2.092±0.463	1.982±0.486
DTLZ5(2,50)	638±101	5681±789	22363±2620	3.168±0.083	1.943±0.263	1.630±0.252
DTLZ5(3,05)	329±50	2444±220	9813±674	1.479±0.048	0.986±0.036	0.893±0.029
DTLZ5(3,20)	238±35	3020±416	14784±1090	3.580±0.056	2.902±0.169	2.793±0.147
DTLZ5(5,10)	219±27	2191±560	12435±1166	3.430±0.054	3.246±0.112	3.040±0.085
DTLZ5(5,20)	290±56	3735±357	14366±983	3.703±0.039	3.142±0.052	2.953±0.039
DTLZ5(7,10)	250±48	1636±387	11919±1324	3.475±0.038	3.387±0.050	3.141±0.083
DTLZ5(7,20)	306±57	1877±986	15687±1482	3.699±0.039	3.569±0.147	3.030±0.058

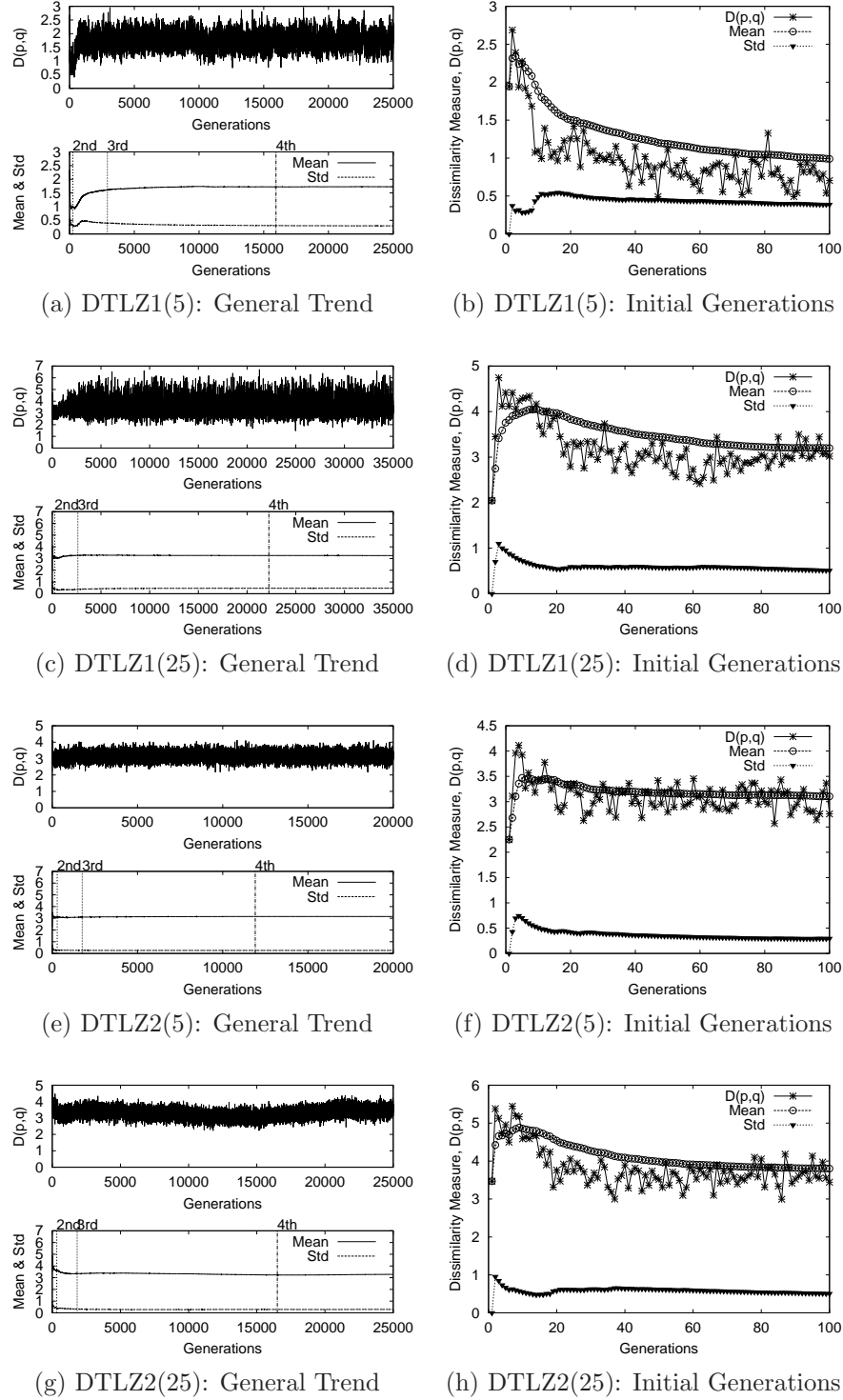


Figure 7.13: DTLZ1( $M$ ) and DTLZ2( $M$ ) where  $M = 5, 25$ : Performance evaluation of Algorithm 6 through the  $\mathcal{D}(p, q)$  measure along with its mean and standard deviation, and a zoomed snapshot limited to first 100 generations.

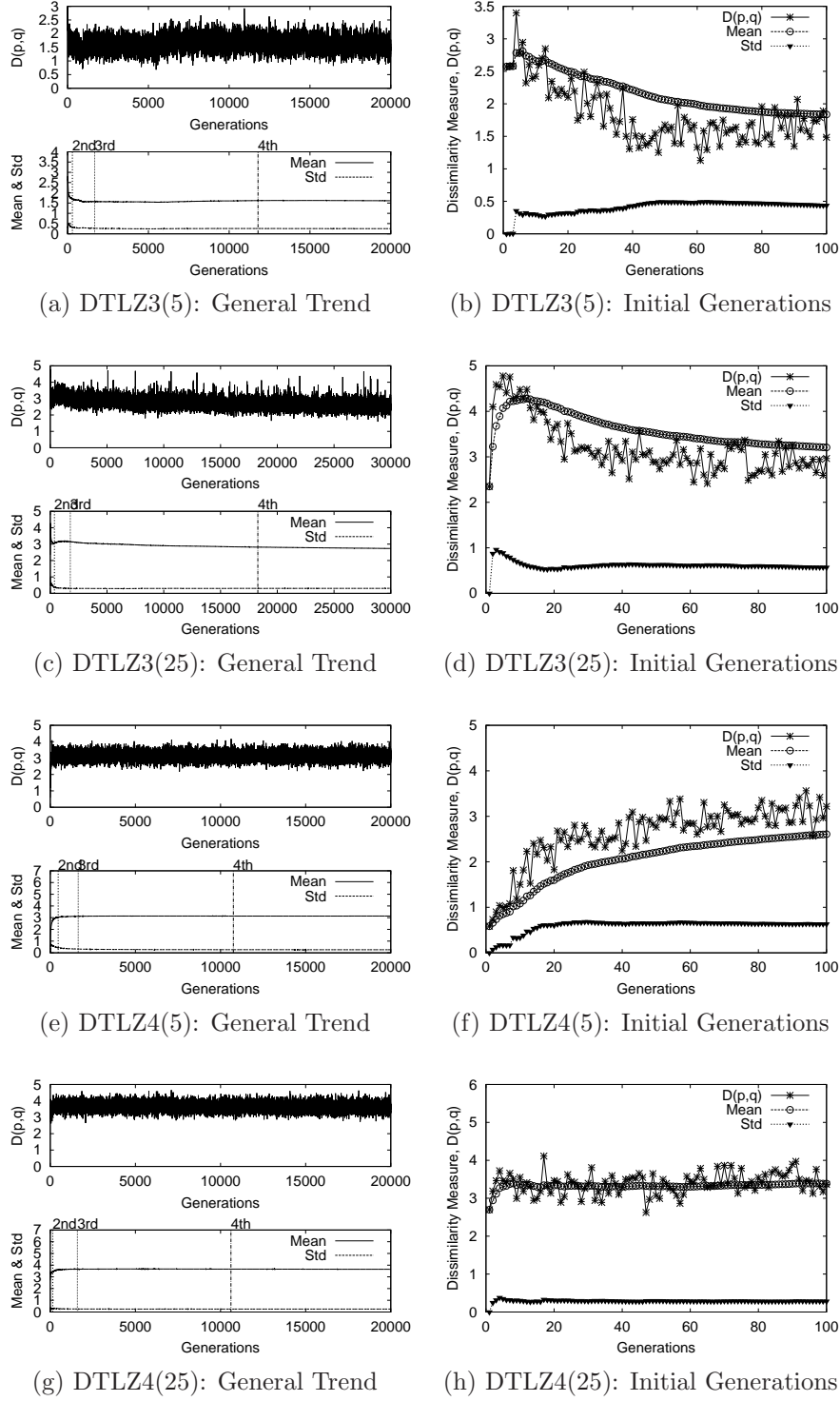


Figure 7.14: DTLZ3( $M$ ) and DTLZ4( $M$ ) where  $M = 5, 25$ : Performance evaluation of Algorithm 6 through the  $\mathcal{D}(p, q)$  measure along with its mean and standard deviation, and a zoomed snapshot limited to first 100 generations.

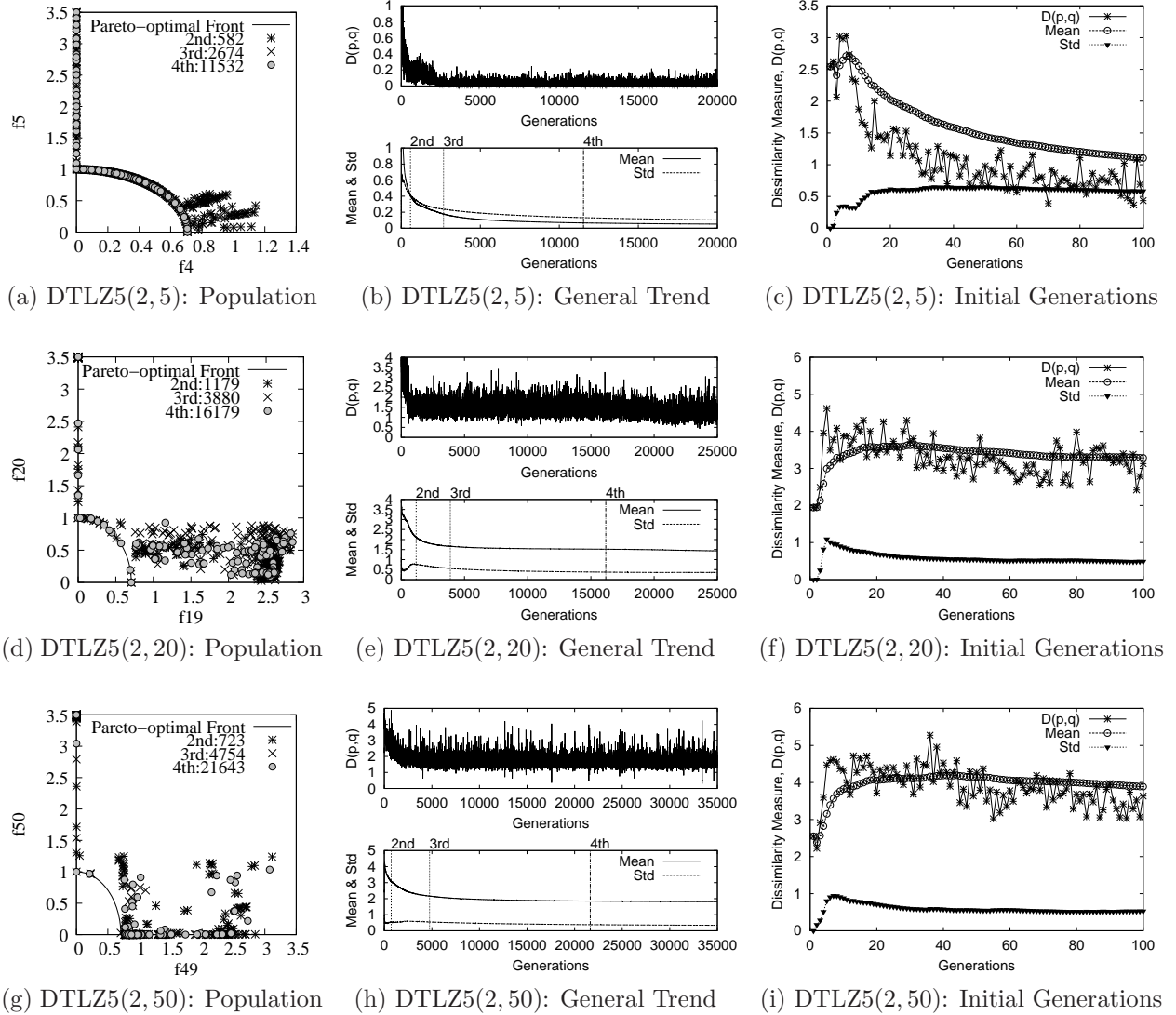
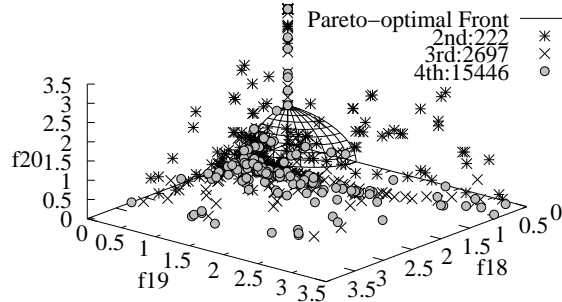
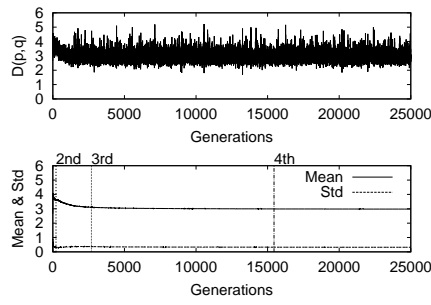


Figure 7.15: DTLZ5(2, 5), DTLZ5(2, 20) and DTLZ5(2, 50): Performance evaluation of Algorithm 6 through the NSGA-II populations at  $N_{gt}$  corresponding to  $n_p = 2, 3, 4$ ; and  $\mathcal{D}(p, q)$  measure along with its mean and standard deviation, and a zoomed snapshot limited to the first 100 generations.

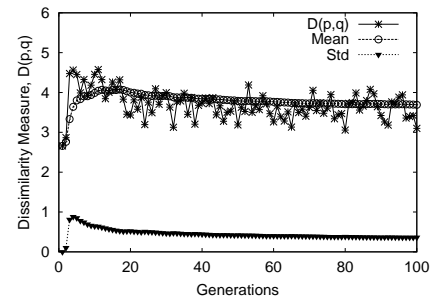




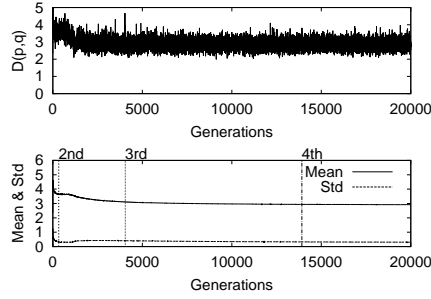
(a) DTLZ5(3, 20): Population



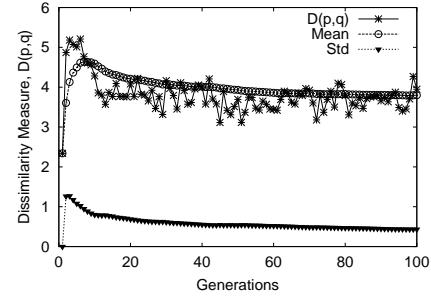
(b) DTLZ5(3, 20): General Trend



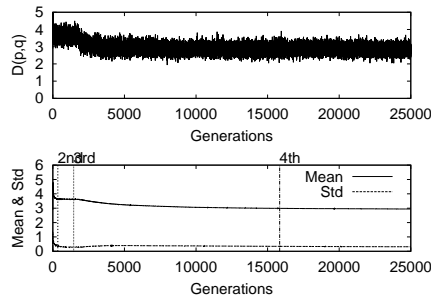
(c) DTLZ5(3, 20): Initial Generations



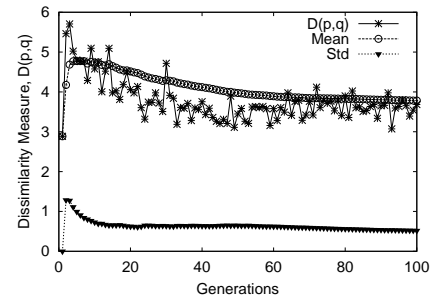
(d) DTLZ5(5, 20): General Trend



(e) DTLZ5(5, 20): Initial Generations



(f) DTLZ5(7, 20): General Trend



(g) DTLZ5(7, 20): Initial Generations

Figure 7.16: DTLZ5(3, 20), DTLZ5(5, 20) and DTLZ5(7, 20): Performance evaluation of Algorithm 6 through the  $\mathcal{D}(p, q)$  measure along with its mean and standard deviation, and a zoomed snapshot limited to first 100 generations. For the particular case of DTLZ5(3, 20), the NSGA-II populations at  $N_{gt}$  corresponding to  $n_p = 2, 3, 4$  are also shown.

may largely occupy different cells, and (iii) the *dissimilarity* measure may be defined only by the non-intersection set, implying high values (away from zero) for the  $\mathcal{D}(p, q)$  measure.

The results presented in Table 7.3 show that:

1. As  $M$  increases, so does the *dissimilarity* measure  $\mathcal{D}(p, q)$  (assuming that the dimension of the POF is the same as the number of objectives).
2. The stagnation (or in contrast—the search efficiency) of NSGA-II is also influenced by the dimensionality of the POF, and is not governed by just the number of objectives. In other words, within the same class of problems, the  $\mathcal{D}(p, q)$  may be smaller even when the number of objectives are higher, provided the dimensionality of the POF is smaller. For instance, in DTLZ5(2, 20) the number of objectives are higher than those in DTLZ5(5, 10), yet the  $\mathcal{D}(p, q)$  value of the former is lower than that of the latter.
3. In a given class of problems, for  $M \geq 10$ , the  $\mathcal{D}(p, q)$  values are quite close to each other. For instance, the mean  $\mathcal{D}(p, q)$  values for: (i) DTLZ4(15) and DTLZ4(25) are 3.5999 and 3.520, respectively, corresponding to  $n_p = 2$ ; and 3.621 and 3.652, respectively, corresponding to  $n_p = 4$ , and (ii) DTLZ5(7, 10) and DTLZ5(7, 20) are 3.475 and 3.699, respectively, corresponding to  $n_p = 2$ , and 3.141 and 3.030, respectively, corresponding to  $n_p = 4$ . It implies, that within the same class of problems, when the number of objectives increased beyond a certain number, the variability of NSGA-II solutions in successive generations tends to be of similar degree.
4. Within a class of problems, in many instances NSGA-II can be seen to stagnate faster when the number of objectives are higher. For instance, corresponding to the case of  $n_p = 2$ , the  $N_{gt}$  for: (i) DTLZ1(5), DTLZ1(15), and DTLZ1(25) is  $357 \pm 158$ ,  $318 \pm 75$ ,  $322 \pm 54$ , respectively, (ii) DTLZ3(5) and DTLZ3(15) is  $336 \pm 100$  and  $282 \pm 40$ , respectively. This could be related to the fact that as the number of objectives increase, more and more solutions tend to be non-dominated from the initial generations itself

due to which the dominance based primary selection becomes ineffective, resulting in poor convergence.

The fact that these experimental results are in line with the expected trends discussed earlier in this chapter (about the  $\mathcal{D}(p, q)$  values) or the previous ones (about the convergence characteristics of density based MOEAs in general, and NSGA-II in particular) add credibility to the performance of the proposed Algorithm 6.

### 7.7.3 Real-World Problems

The issue of how long an MOEA needs to be run becomes crucial in the case of real-world problems where nothing about the Pareto-optimal solutions may be known *a priori*. If the pre-fixed  $N_g$  is too high and the problem is solved in too few generations it would imply waste of computational resources. On the contrary, if the pre-fixed  $N_g$  is insufficient, it would result in a poor POF-approximation.

In the case of real-world MaOPs, where most existing MOEAs are likely to provide a good POF-approximation, the application of the objective reduction framework holds promise, as it may potentially reveal the presence of some redundant objectives, and the relative preferences of the different objectives in constituting the POF. The rationale for the choice of objective reduction based on preservation of the correlation-structure of a given non-dominated solution set has already been discussed earlier in this thesis. This thesis has also: (i) introduced the notion of a good POF-representation by a given solution set, which implies that its correlation-structure is strongly in conformance with that of the true POF, (ii) discussed that the success of the proposed objective reduction framework relies on a good POF-representation by the given solution set, and (iii) identified that a good POF-approximation is a *sufficient* but not a *necessary* condition for a good POF-representation. It is in this context that it becomes critical to determine whether the MOEA has:

- converged to the POF, i.e., a good POF-approximation has been obtained.

Table 7.4: Highlighting the utility of the proposed MOEA-termination algorithm (Algorithm 6) for real-world problems.

(a) Left hand side table quotes the results obtained in Chapter 5 corresponding to  $N_g = 2000$  for NSGA-II. The right hand side table presents the  $N_{gt}$  and  $\mathcal{D}(p, q)$  values determined by Algorithm 6 along an NSGA-II run.

Problem	Objective Preference Information		Algorithm 6 results					
	Essential objective set	Preference-ranking of all objectives	$N_{gt}$			$\mathcal{D}(p, q)$		
	$(\mathcal{F}_s)$	( <i>more-important-than</i> relation)	$n_p=2$	$n_p=3$	$n_p=4$	$n_p=2$	$n_p=3$	$n_p=4$
Car Side-impact	$\{f_1, f_6, f_7, f_8, f_9, f_{11}\}$	$f_1 \triangleright f_7 \triangleright f_6 \triangleright f_8 \triangleright f_{11} \triangleright f_9 \triangleright f_5 \triangleright f_2 \triangleright f_3 \triangleright f_4 \triangleright f_{10}$	292	1477	10758	3.19234	3.83883	3.25857
Storm Drainage	$\{f_2, f_3, f_4, f_5\}$	$f_4 \triangleright f_3 \triangleright f_2 \triangleright f_5 \triangleright f_1$	228	1103	9348	1.57972	1.49190	1.58658
Work Roll Cooling	$\{f_1, f_4\}$	$f_4 \triangleright f_1 \triangleright f_2 \triangleright f_5 \triangleright f_3 \triangleright f_6$	224	950	7734	1.19008	1.33320	1.31208

(b) Objective Reduction Framework results corresponding to  $N_g = 1$  and  $N_g = N_{gt}$ , where  $N_{gt}$  is determined by Algorithm 6

Problem	Objective Reduction Framework results corresponding to $N_{gt}$ determined by Algorithm 6			
	$N_g = 1$	$n_p=2$	$n_p=3$	$n_p=4$
Car Side-impact	$\mathcal{F}_s = \{f_1, f_5, f_7, f_8, f_{11}\}$ $f_1 \triangleright f_8 \triangleright f_7 \triangleright f_{11} \triangleright f_5 \triangleright f_9 \triangleright f_6 \triangleright f_2 \triangleright f_3 \triangleright f_4 \triangleright f_{10}$	$\mathcal{F}_s = \{f_1, f_2, f_3, f_6, f_7, f_8, f_9, f_{11}\}$ $f_1 \triangleright f_7 \triangleright f_6 \triangleright f_8 \triangleright f_{11} \triangleright f_9 \triangleright f_2 \triangleright f_3 \triangleright f_5 \triangleright f_4 \triangleright f_{10}$	$\mathcal{F}_s = \{f_1, f_6, f_7, f_8, f_9, f_{11}\}$ $f_1 \triangleright f_7 \triangleright f_6 \triangleright f_8 \triangleright f_{11} \triangleright f_9 \triangleright f_5 \triangleright f_2 \triangleright f_3 \triangleright f_4 \triangleright f_{10}$	$\mathcal{F}_s = \{f_1, f_6, f_7, f_8, f_9, f_{11}\}$ $f_1 \triangleright f_7 \triangleright f_6 \triangleright f_8 \triangleright f_{11} \triangleright f_9 \triangleright f_5 \triangleright f_2 \triangleright f_3 \triangleright f_4 \triangleright f_{10}$
Storm Drainage	$\mathcal{F}_s = \{f_2, f_3, f_4\}$ $f_4 \triangleright f_3 \triangleright f_2 \triangleright f_5 \triangleright f_1$	$\mathcal{F}_s = \{f_2, f_3, f_4, f_5\}$ $f_4 \triangleright f_3 \triangleright f_2 \triangleright f_5 \triangleright f_1$	$\mathcal{F}_s = \{f_2, f_3, f_4, f_5\}$ $f_4 \triangleright f_3 \triangleright f_2 \triangleright f_5 \triangleright f_1$	$\mathcal{F}_s = \{f_2, f_3, f_4, f_5\}$ $f_4 \triangleright f_3 \triangleright f_2 \triangleright f_5 \triangleright f_1$
Work Roll Cooling	$\mathcal{F}_s = \{f_1, f_4\}$ $f_4 \triangleright f_1 \triangleright f_2 \triangleright f_5 \triangleright f_3 \triangleright f_6$	$\mathcal{F}_s = \{f_1, f_4\}$ $f_4 \triangleright f_1 \triangleright f_2 \triangleright f_5 \triangleright f_3 \triangleright f_6$	$\mathcal{F}_s = \{f_1, f_4\}$ $f_4 \triangleright f_1 \triangleright f_2 \triangleright f_5 \triangleright f_3 \triangleright f_6$	$\mathcal{F}_s = \{f_1, f_4\}$ $f_4 \triangleright f_1 \triangleright f_2 \triangleright f_5 \triangleright f_3 \triangleright f_6$

- stagnated, implying it can not improve the quality of the POF–approximation, any further.

This is critical because in either of the above cases, running the MOEA any further will offer no further improvement in the POF–approximation, despite additional computational expense.

This section demonstrates as to how the proposed Algorithm 6 helps to determine as to when any one of the above two conditions could be ascertained, so that the underlying MOEA could be terminated and the objective reduction framework be applied. This is achieved by:

- re-visiting the real-world problems considered in the previous chapters, where the objective reduction framework was applied for an arbitrarily fixed  $N_g = 2000$  for NSGA-II (Table 7.4a).
- comparing the previous results with the newly generated results (by applying the objective reduction framework) corresponding to  $N_{gt}$  determined by Algorithm 6 (Table 7.4b). It needs to be noted that while generating the new results, no changes have been made to the population size ( $N = 200$ ) or the NSGA-II related parameters.

The problems being revisited here have been discussed in Chapter 5 and include the 11-objective car side-impact problem, the five-objective storm drainage problem, and the six-objective work roll cooling design problem.

From the results summarised in Table 7.4, the following observations can be made:

- that, regardless of the chosen  $n_p$ , the  $\mathcal{D}(p, q)$  values are quite high (far from zero), for each of the three problems (Table 7.4a). This indicates that NSGA-II has failed to converge to the POF, corresponding to the  $N_{gt}$  determined by Algorithm 6, where  $N_{gt}$  ranges from values as low as 224 (in work roll cooling problem:  $n_p = 2$ ) to 10758 (in car side-impact problem:  $n_p = 4$ ).

- the results obtained from the objective reduction framework corresponding to  $N_g = 1$  do not match with those corresponding to  $N_g = 2000$ , in two out of the three problems (Table 7.4b). This suggests that the correlation-structure of the NSGA-II population at  $N_g = 1$  does not coincide with that corresponding to  $N_g = 2000$ . Notably there is no rationale for pre-fixing  $N_g = 2000$  at first place, except that NSGA-II is provided reasonable computational effort for search.
- barring one exception, the results obtained by the objective reduction framework corresponding to arbitrarily fixed  $N_g = 2000$  and  $N_{gt}$  determined by Algorithm 6 (corresponding to three different  $n_p$ ) coincide. The only exception relates to the car side-impact problem with  $n_p = 2$ .

The above observations demonstrate the utility of the proposed MOEA-termination algorithm (Algorithm 6). For instance:

- the inconsistency between the results corresponding to  $N_g = 1$  and  $N_g = 2000$  highlights the need for some criterion that could help decide a reasonably reliable  $N_g$  (for a given problem and a given MOEA) corresponding to which the objective reduction framework could be applied.
- the consistency of the results at  $N_{gt}$  determined by Algorithm 6 corresponding to  $n_p = 2, 3, 4$  (barring one case highlighted above), provides strong evidence in favour of the argument that the proposed Algorithm 6 provides the desired termination criterion.
- corresponding to  $n_p = 4$ , the  $N_{gt}$  for the car side-impact, storm drainage, and work roll cooling problem is 10758, 9348, and 7734, respectively, each of which is way higher than the earlier pre-fixed  $N_g = 2000$ . Though the application of Algorithm 6 in this case implies higher computational cost towards obtaining the same results (as at  $N_g = 2000$ ), it also provides a rationale as to why the objective reduction framework is being applied after a certain number of generations of NSGA-II.

- Corresponding to  $n_p = 3$ , the  $N_{gt}$  for the car side-impact, storm drainage, and work roll cooling problem is 1477, 1103, and 950, respectively, each of which is less than the earlier pre-fixed  $N_g = 2000$ . This demonstrates the potential of computational savings without compromising on the results.
- Corresponding to  $n_p = 2$ , the  $N_{gt}$  for the car side-impact, storm drainage, and work roll cooling problem is 292, 228, and 224, respectively, each of which is far less than the earlier pre-fixed  $N_g = 2000$ . This again demonstrates the potential of computational savings without compromising much on the results.

Arguably, the benefits of the application of Algorithm 6 highlighted above are specific to the problems considered. Nonetheless, based on the experimental results on a wide variety of test problems and the real-world problems, it can be inferred that the  $N_{gt}$  determined by Algorithm 6 corresponding to  $n_p = 3$  could be treated with a reasonably high degree of confidence.

## 7.8 Summary

In this chapter, an entropy based *dissimilarity* measure has been proposed, which determines when (in terms of the number of generations of an MOEA) the *sufficiency* condition of a good POF-approximation is met, or that it can not be met. In doing so, it reveals the generation count at which the underlying MOEA could be terminated. The low computational complexity of the proposed measure makes it applicable to any multi-objective problem (regardless of the number of objectives) and its generic nature makes it easily integrable with existing MOEAs. Experimental results have revealed that: (i) in the case of two- and three-objective problems, the proposed *dissimilarity* measure was able to indicate when NSGA-II could converge to the POF (subject to user-defined precision levels), and (ii) in the case of MaOPs, it was able to indicate the stagnation of NSGA-II, i.e., its inability to improve upon the quality of the POF-approximation beyond a certain level.

## Chapter 8

# Discussion and Conclusions

MaOPs are known to pose difficulties in terms of the search efficiency of the underlying MOEA, computational cost, decision making and visualisation. Recognising that the complete POF-approximation of MaOPs with reasonable computational resources is very difficult, there is a growing interest towards employing the MCDM based MOEAs to tackle such problems. In that, the focus is on utilising the DMs' preferences—*a priori* or interactively, and guiding the search towards a chosen few solutions, rather than the whole POF. While this approach promises to counter some of the difficulties associated with MaOPs, it has its own pitfalls. Some of these pitfalls relate to the fact that in real-world problems, a single DM rarely exists, and multiple DMs may have conflicting preferences based on their individual experience, memory, thoughts, thinking paradigms or psychological states. Furthermore, research in psychology and cognitive sciences has revealed that people are neither very good at handling large amount of data nor expressing them in quantified ways, nor at handling more than several  $(7 \pm 2)$  factors at the same time. Given these factors, the DMs' preferences may result in lack of *objectivity* (a rational basis); *repeatability* (identical preferences for identical options); *consistency* (alike preferences across multiple interaction stages); and *coherence* (alike preferences by multiple DMs).

This thesis is based on the premise that the pitfalls associated with the MCDM based



MOEAs could be overcome by *learning* the preference-structure of the different objectives inherent in the problem model itself, and utilising it towards a decision support for the DMs. The justification for this premise lies in the fact that most often, the problem models are an outcome of the joint expertise of the engineers and technical managers. Hence the preference-structure inherent in these models should be utilised to support the DMs in articulating their preferences with rationality. Notably, this approach holds promise for a class of MaOPs where it may turn out that not all the objectives are *essential* to describe the POF, or not all the objectives are equally important. In this background, this thesis aimed at developing a framework for both *offline* and *online* decision support, that promises to facilitate the DMs preferences with *objectivity*, *repeatability*, *consistency*, and *coherence*.

## 8.1 Discussion

The aim of this thesis has been realised through the following objectives:

1. identification of a meaningful criterion for developing the decision support.
2. development of a machine learning based framework towards the envisioned (*offline*) decision support in terms of:

- I revelation of an *essential* objective set—the smallest set of conflicting objectives which can generate the same POF as the one obtained by the original problem.
- II preference-ranking of all the objectives in terms of a preference-weight for each objective that varies from zero to one, and the sum of all weights is one.
- III revelation of the smallest objective set—retaining which will ensure that the error incurred is *just less than or equal to* an error specified by the DMs.
- IV revelation of the objective sets of sizes specified by the DMs, such that these sets correspond to minimal error.
- V visual representation of the above revelations in one snap-shot.

3. development of an *online* implementation that counters some of the pitfalls associated with the *offline* implementation of the framework.
4. addressal of the fundamental question of when to *time* the decision support.
5. experimental validation based on a wide range of test problems and some real-world problems.

### 8.1.1 Contribution to Knowledge

The main contribution to knowledge of this thesis is the proposal of a machine learning based framework that offers both *offline* and *online* decision support, facilitating *objectivity*, *repeatability*, *consistency* and *coherence* in the DMs' preferences, when dealing with MaOPs. This contribution lies in its:

1. recognition that some of the difficulties associated with MaOPs could be countered by integrating the MCDM approaches with MOEAs. In that, the DMs' preferences could be utilised to guide the MOEA search towards the preferred solutions, as opposed to the whole POF.
2. recognition that in practical scenarios: (i) multiple DMs may be involved with possibly conflicting preferences, (ii) the DMs may need to rationally justify the preferences made, specially in mission critical applications where the security, enormous cost or human aspects are involved, and (iii) owing to the cognitive limitations, the DMs may not be in a position to articulate their preferences with a rational basis, and hence, these preferences may not be *repeatable*, *consistent*, and *coherent*.
3. recognition that the above pitfalls associated with the DMs' preferences could be countered, in principle, by utilising the preference-structure of the objective functions inherent in the problem model developed by the domain experts.

4. recognition that in a given problem, not all the objectives may be *essential* to describe the POF, or not all the objectives may be equally important; and, that this information could be derived by applying machine learning techniques to the best non-dominated solutions obtained from an MOEA corresponding to the given problem model.

In this process, this thesis has:

1. demonstrated that an MOEA such as NSGA-II which is well known for its efficacy in approximating the POF for two- and three-objective problems, provides a poor POF-approximation (convergence to and diversity across the POF) in the case of MaOPs<sup>1</sup>. In that, the dominance relations characterising the approximated (non-dominated) solution set may be different from those characterising the true POF—a feature that is referred in this thesis, as *noise*.
2. revealed that the presence of redundant objectives hinders the search efficiency of an MOEA, resulting in a poor POF-approximation.
3. revealed that even when the dominance relations characterising the approximated solution set and the true POF may be different, their global correlation structure may still be similar. In this regard, a distinction has been made between a good POF-approximation and a good POF-representation (good conformance between the correlation structure of a solution set and the POF), and it is highlighted that a good POF-approximation is a *sufficient but not a necessary condition* for a good POF-representation, implying:

- (a) if the approximated solution set provides a good POF-approximation, then by *sufficiency* condition, it will also provide a good POF-representation.

---

<sup>1</sup>These observations about NSGA-II could be generalised to most existing MOEAs, and for other density based MOEAs (such as SPEA2 and PESA) in particular. This is because, while dealing with MaOPs, the entire MOEA population acquires the same-rank of non-domination, given which the Pareto-dominance based primary selection is ineffective. Towards the secondary selection based on diversity preservation, the density based MOEAs worsen the situation by favouring the remote and boundary solutions, implying that the best diversity gets associated with poorly converged solutions.

- (b) if the approximated solution set fails to provide a good POF-approximation, then it *not being a necessary* condition, it could still provide a good POF-representation.

In other words, the requirement of a good POF-representation is less stringent than that of a good POF-approximation, hence, if inferences about the true POF are to be drawn based on the approximated solution sets that are typically *noisy* in the case of MaOPs, then preservation of the global correlation-structure is a more promising criterion than preservation of dominance relations.

4. proposed a machine learning based *offline* framework that operates on the objective vectors of the approximated solution set (non-dominated solutions obtained from an MOEA); *learns* the preference-structure of the objective functions by preserving the correlation-structure of the solutions; and provides the decision support in terms of:

- I revelation of an *essential* objective set: here, for a given  $M$ -objective problem denoted by  $\mathcal{F}_0 = \{f_1, f_2, \dots, f_M\}$ , the framework reveals an *essential* objective set—the smallest set of conflicting objectives ( $\mathcal{F}_T, |\mathcal{F}_T| = m$  ( $m \leq M$ )) which can generate the same POF as that by the original problem.
- II preference-ranking of all the objective functions: here, the framework identifies the preference-weight ( $w_i$ ) for each objective, such that  $w_i \geq 0$  and  $\sum_{i=1}^M w_i = 1$ .
- III  $\delta$ -MOSS ( $\delta$ -Minimum Objective Subset) analysis: this analysis imitates a situation where the DM allows for a  $\delta$  ( $0 \leq \delta \leq 1$ ) error, and wants to know the smallest objective set—retaining which would ensure that the error incurred is *just less than or equal to*  $\delta$ . Such an objective set is referred to as the  $\delta$ -minimal set.
- IV  $k$ -EMOSS (Minimum Objective Subset of Size  $k$  with Minimum Error) analysis: this analysis imitates a situation where the DM specifies the fraction  $p$  ( $0 \leq p \leq 1$ ) of the original number of objectives ( $M$ ) to be retained, and wants to know an objective set of size  $k = \lceil pM \rceil$  that corresponds to minimal error. Such an objective set is referred to as the  $k$ -minimal set.

V visual representation: a simple yet meaningful visual representation of the above analysis is presented that could serve as a snap-shot guide for the DM to base his or her preferences on.

This framework comprises of a principal component analysis (PCA) based linear objective reduction algorithm, namely, L-PCA, and a maximum variance unfolding (MVU) based nonlinear objective reduction algorithm, namely, NL-MVU-PCA. In that:

- (a) the L-PCA applies PCA directly in the objective space—represented by the correlation matrix ( $R$ ) that is computed by the objective vectors of the non-dominated solutions obtained from an MOEA. Its scope is limited to linear objective reduction, as PCA is based only on removing the *second order dependencies* (i.e., diagonalisation of  $R$ ), and hence it may be ineffective in dealing with data sets with multi-modal Gaussian or non-Gaussian distributions.
- (b) the NL-MVU-PCA applies PCA in the kernel space—represented by the kernel matrix ( $K$ ). The kernel space is the objective space that has been *unfolded* using the MVU principle, i.e., the space in which the *higher order dependencies* have been removed. Mathematically, the *unfolding* is posed as a semi-definite programming (SDP) problem, the output of which is  $K$ . With the application of PCA in the kernel space (i.e., diagonalisation of  $K$ ), the limitations of L-PCA are countered.

Notably, the framework *learns* the preference-structure of the objective functions by eliminating the redundant objectives—those that are: (i) either non-conflicting along the *significant* eigenvectors of  $R$  in the case of L-PCA, and  $K$  in the case of NL-MVU-PCA, or (ii) are *globally correlated*. The framework is designed to operate iteratively, where the objectives identified as redundant in an iteration are not included in the subsequent iteration, and it terminates when no further reduction of objective(s) in an iteration is possible.

5. incorporated *de-noising* as a central feature of the proposed framework. In the case of

DRP based algorithms, the presence of even a single *noisy* solution leads to inaccurate inferences about the preference-structure of the objective functions on the true POF. To counter this pitfall, the framework (both the L-PCA and NL-MVU-PCA) proposed in this thesis has *de-noising* mechanisms integrated in each of its objective reduction component, as follows:

- (a) while eliminating the objectives that are non-conflicting along the *significant* eigenvectors, the *de-noising* is implicitly performed by not accounting for the potentially *weak/noisy* inter-relationships between objectives captured by the *insignificant* eigenvectors.
- (b) while eliminating the objectives that are *globally correlated*, the *de-noising* is performed by testing the strength of correlation between two potentially correlated objectives (those whose signs of correlation with all other objectives are identical) against a correlation threshold. In that, two potentially correlated objectives are considered as globally correlated only if the strength of their correlation is higher than a threshold. The notion of threshold is introduced to account for the fact that due to the presence of some *noisy* solutions, two conflicting objectives may inaccurately exhibit global correlation<sup>2</sup>. A commonly referred criterion for interpreting the strength of correlation, is the Cohen scale, which refers a correlation strength of 0.1 to 0.3 as *weak*, 0.3 to 0.5 as *moderate* and 0.5 to 1.0 as *strong*. However, in the context of the objective reduction framework, it is desired that the correlation threshold be *dynamically* determined depending upon the nature of the underlying problem. In that, a high correlation threshold is desired for problems that have none or only a few redundant objectives (thereby, prohibiting/restricting objective reduction based on global correlation), while a low correlation threshold

---

<sup>2</sup>It may also happen that two globally correlated objectives may exhibit conflict. In such a case, no objective could be eliminated in that iteration. However, in the subsequent iteration where the number of objectives may be fewer, approximation sets with a better POF-representation can be expected, enabling further objective reduction based on global correlation.

is desired for problems that have high redundancy (thereby, facilitating objective reduction based on global correlation). In this thesis, a dynamic computation of the correlation threshold based on the spectrum of the eigenvalues (for a given problem) has been proposed.

6. accomplished the framework—constituted by the L-PCA and NL-MVU-PCA with computational complexities [177] corresponding to  $O(NM^2 + M^3)$  and  $O(M^6)$  (worst case), respectively<sup>3</sup>. Notably, these algorithms have a higher computational efficiency than the alternative algorithms (well reported in literature) available for generating the perceived decision support. For example, two versions of the DRP based algorithms exist, namely: (i) the exact algorithm which guarantees a *0-minimum* objective set<sup>4</sup>, and (ii) the greedy algorithm that does not guarantee a *0-minimum* objective set. The computational complexities of the exact and greedy algorithms correspond to  $O(N^2M2^M)$  and  $O(\min\{N^2M^3, N^4M^2\})$ , respectively. In the case of MaOPs,  $N \gg M$  as the requirement of  $N$  grows exponentially with  $M$ , and hence, arguably: (i) the DRP based exact algorithm becomes rather impractical to use, since it is exponential in  $M$  and quadratic in  $N$ , and (ii) even the greedy algorithm is likely to be computationally more expensive than NL-MVU-PCA (worst case being  $O(M^6)$ ).
7. proposed an *online* implementation of the framework. The *offline* objective reduction framework discussed above operates in an iterative manner, in that, only the objectives identified as *essential* in a particular iteration are considered in the subsequent iteration. For solution sets with a poor POF-representation, it is likely that an *essential* objective may erroneously be eliminated as redundant, with no scope of being reconsidered. This limitation has been overcome in an *online* implementation of the framework, where all the objectives are retained with different probabilities, at a given instant. This implementation is able to enhance the accuracy of the framework given

---

<sup>3</sup>Recall, that  $M$  represents the number of objectives, and  $N$  represents the number of solutions.

<sup>4</sup>*0-minimum* objective set implies solution to the  $\delta$ -MOSS problem with  $\delta = 0$ , and corresponds to an *essential* objective set for a given problem.

that the computational effort spent in an iteration can be exploited in subsequent iterations, as opposed to the *offline* implementation (where the population for a subsequent iteration needs to be re-initialised).

8. addressed the fundamental question as to when the proposed framework should be applied. Towards it, this thesis proposed an entropy based *dissimilarity* measure, which along an MOEA run identifies one of the following characteristics (as the case may be for a given problem, and a given MOEA):

- (a) whether or not the MOEA has converged to the POF, i.e., whether or not a good POF-approximation has been obtained.
- (b) whether or not the MOEA has stagnated, in terms of its inability to further improve the POF-approximation, despite additional computational expense.

Given that the proposed framework relies on a good POF-representation, for which a good POF-approximation is a *sufficient* condition, it can be inferred that the purposed *dissimilarity* measure identifies whether the condition of a good POF-approximation is met, or that it can not be met. In either case, there is no advantage of running the underlying MOEA any further. While it provides a termination criterion for MOEAs in general, in the particular case of MaOPs (where Item b, is more likely), it determines as to when the underlying MOEA be terminated and the proposed machine learning based framework be applied. The generic nature (applicability to all multi-objective problems, regardless of the number of the objectives) of the *dissimilarity* measure, and its low computational complexity is a significant contribution of this thesis. Experiments on an extensive set of test problems have revealed that in the case of two- and three-objective problems, the measure is able to accurately indicate whether an MOEA has converged to the POF or not, and in the case of MaOPs, it is able to indicate whether an MOEA has stagnated, in terms of its inability to further improve the quality of the POF-approximation.



### 8.1.2 Experimental Validation and Key Findings

The proposed framework has been tested against a broad range of test problems and also some real-world MaOPs. A chapter summary of the experiments conducted and the corresponding key findings, are summarised below.

- In Chapter 4, the results correspond to over 9000 simulations performed on 24 versions of six test problems and two real-world problems. In that, the performance of the proposed L-PCA and NL-MVU-PCA is tested against the DRP based algorithms, corresponding to the solutions generated—on the POF ( $\mathcal{N}_{\mathcal{P}}$ ); randomly ( $\mathcal{N}_{\mathcal{R}}$ ); by 20 runs each of NSGA-II ( $\mathcal{N}_{\mathcal{NS}}$ ) and  $\epsilon$ -MOEA ( $\mathcal{N}_{\epsilon}$ ). The sensitivity of the proposed algorithms on randomness; the critical parameter involved; and the characteristics of the underlying non-dominated sets, is also discussed. The key findings based on these simulations are as follows:
  - The DRP based algorithms are extremely sensitive to *noise*, and unlike in the case of  $\mathcal{N}_{\mathcal{P}}$ , these algorithms largely failed to accurately identify an *essential* objective set for the other data sets. In the case of MaOPs, as the solution sets obtained from most existing MOEAs are likely to be characterised by *noise*, the merit of preserving the dominance relations is questionable, at first place.
  - While both  $\mathcal{N}_{\mathcal{NS}}$  and  $\mathcal{N}_{\epsilon}$  failed to provide a good POF–approximation, they provided a reasonably good POF–representation<sup>5</sup>. This explains the significantly better (compared to the DRP based algorithms) performance of the proposed L-PCA and NL-MVU-PCA. In some instances where the POF–representation was not good:
    - \* in the sense that not all the correlated objectives were captured by the solution set, these algorithms could identify an *essential* objective set through the

---

<sup>5</sup>Recall that a good POF–approximation is a *sufficient* but not a *necessary* condition for a good POF–representation.

iterative approach. In that: (i) the objectives that were accurately captured as correlated, were eliminated in that iteration, and (ii) in the subsequent iteration where fewer objectives were involved, a better POF-representation was obtained and the remaining correlated objectives were eliminated, leading to an *essential* objective set.

- \* in the sense that even some conflicting objectives (on the true POF) were found as correlated, these algorithms through their *de-noising* feature based on the correlation threshold could disregard the misleading correlations, accurately leading to an *essential* objective set.

Between the proposed algorithms, NL-MVU-PCA performs better than L-PCA, owing to the fact that it applies PCA in the kernel space, i.e., the nonlinearly *unfolded* objective space, and that allows for accurate determination of the principal components.

- In Chapter 5, the results correspond to over 4000 simulations performed on 12 versions of two test problems and four real-world problems. In that, the  $\delta$ -MOSS and  $k$ -EMOSS analysis is conducted, and the performance of NL-MVU-PCA is tested against the DRP based exact algorithm<sup>6</sup>. For one of the real-world problems, the solution set corresponding to MSOPS-II ([43]) is used, while for the remaining ones 20 different sets of  $\mathcal{N}_{\mathcal{N}\mathcal{S}}$  are used. For the test problems, the *noise* free  $\mathcal{N}_{\mathcal{P}}$ , and 20 different sets of  $\mathcal{N}_{\epsilon}$  are used. The key findings based on these simulations are as follows:
  - Corresponding to  $\mathcal{N}_{\mathcal{P}}$ , both the NL-MVU-PCA and DRP based exact algorithm could accurately identify the  $\delta$ -minimal set for  $\delta = 0$ ; and the  $k$ -minimal error corresponding to: (i)  $k = I$  in the case of DTLZ5( $I, M$ ), and (ii)  $k = M$  in the case of non-redundant problems.

---

<sup>6</sup>The L-PCA and the DRP based greedy algorithm are not considered, based on their inferior performance, reported in Chapter 4.

- Corresponding to  $\mathcal{N}_{\mathcal{P}}$ , the DRP based exact algorithm failed to capture the problem features through the  $\delta$ -MOSS analysis with  $\delta > 0$ , and  $k$ -EMOSS analysis with  $k < M$ . For instance, in a non-redundant problem like DTLZ2<sub>BZ</sub>, each objective is in conflict with the rest, and hence:

- \* in the case  $\delta$ -MOSS analysis, as  $\delta$  gradually increases from 0 to 1,  $|\mathcal{F}_{\{\delta\}}|_s$  should gradually decrease.
- \* in the case of  $k$ -EMOSS analysis, as  $k$  gradually decreases from  $M$  to 1,  $\mathcal{E}_k^n$  should gradually increase.

Notably, while the NL-MVU-PCA captures the expected trend, the DRP based exact algorithm fails. Given that the underlying data set, namely  $\mathcal{N}_{\mathcal{P}}$ , is free of *noise*, the failure of the exact algorithm could be attributed to the stringent assumptions on which the definition of  $\delta$  error is based.

- Corresponding to  $\mathcal{N}_{\epsilon}$ , it is observed that while NL-MVU-PCA largely captures the problem features through the  $\delta$ -MOSS and  $k$ -EMOSS analysis, the DRP based exact algorithm largely fails. The latter could be jointly attributed to the stringent assumptions on which it is based, and its sensitivity to *noise*.
- In Chapter 6, the results correspond to over 480 simulations performed on 9 different versions of one test problem, and 3 different versions of another test problem. For test problems, the simulations correspond to 20 different sets each, of  $\mathcal{N}_{\epsilon-p-On}$  (solutions generated by  $\epsilon$ -MOEA, corresponding to objectives chosen by the Framework 4) and  $\mathcal{N}_{\epsilon-Off}$  (solutions generated by  $\epsilon$ -MOEA, corresponding to objectives chosen by the Framework 1). The key finding based on these simulations is that: (i) the probabilistic retention of all the objectives by the Framework 4 serves as a self-correcting mechanism in the wake of erroneous preference information by objective reduction algorithms, and (ii) while the above promises enhanced accuracy, it does so at the cost of higher computational effort.

- In Chapter 7, the results correspond to over 3960 simulations performed on 22 two-objective problems, one three-objective problem, five MaOPs and three real-world problems. For each problem, the simulations correspond to 10 different sets of solutions obtained from NSGA-II, and three different settings of the algorithm parameter  $n_p$  that defines the degree of accuracy for the termination criterion. The key findings are as follows:
  - For two- and three-objective problems, the proposed *dissimilarity* measure is successfully able to identify when the underlying MOEA converges to the POF.
  - In the case of MaOPs, including both the test problems and the real-world problems, the proposed *dissimilarity* measure is successfully able to identify when the underlying MOEA population has stabilised—implying that it has stagnated away from the POF. In doing so, the proposed measure is able to reveal as to when the machine learning based framework be *timed* (applied), and the most meaningful decision support possible be offered.
  - The other termination criterion algorithms available in the literature often employ more than one distance measure. However, in the algorithm proposed in this thesis, the underlying *dissimilarity* measure alone is sufficient to identify as to when an MOEA population stabilises.

## 8.2 Conclusions

The objectives pursued by this thesis have been achieved as follows:

1. identification of a meaningful criterion for developing a decision support: it has been identified in Chapter 4 that a criterion based on preserving the correlation–structure of the given data set is more meaningful than preserving the dominance relations. This has been shown for a wide range of test problems with different degrees of redundancy and *noise*, and for two real-world problems.

2. development of a machine learning based framework towards the envisioned decision support: a machine learning based framework has been proposed in Chapter 5 that provides a decision support in the following forms: (i) revelation of an essential objective set, (ii) preference-ranking of all the objectives, (iii)  $\delta$ -MOSS and  $k$ -EMOSS analysis and (iv) a meaningful visual representation of the analysis performed by the previous points. This has been shown for a wide range of test problems with different degrees of redundancy and *noise*, and for four real-world problems.
3. to explore the benefits of implementing *online* the proposed framework: an *online* implementation of the proposed framework is provided in Chapter 6. The benefits that have been revealed are: (i) that the probabilistic retention of all the objectives in a particular iteration serves as a self-correcting mechanism, and enables accurate preferences information in subsequent iterations of the framework and (ii) despite higher computational effort the *online* implementation revealed more accuracy when compared with the *offline* implementation. This has been shown for 9 different versions of one test problem and three different versions of another test problem.
4. an address of the fundamental question as to when the proposed framework should be applied: a *dissimilarity* measure has been proposed in Chapter 7 that is able to identify when the MOEA population stabilises. This implies that the proposed measure is capable of revealing when the machine learning based framework should be *timed* (applied) which leads to the most meaningful decision support possible that can be offered. This has been shown for a wide range of test problems and three real-world problems.
5. experimental validation and comparative analysis through a wide range of test problem and real-world problems: this objective has been achieved throughout the experimental results sections conducted in Chapters 4, 5, 6 and 7.

The key findings of this thesis are:

- In Chapter 4, it has been shown that the DRP based algorithms are extremely sensitive to *noise* and therefore these algorithms fail to accurately identify an *essential* objective set for noisy data sets. This implies that the merit of preserving the dominance relations of noisy data sets is itself questionable. The proposed L-PCA and NL-MVU-PCA have reported a significantly better performance when compared with DRP based algorithms. This can be attributed to a reasonably good POF-representation that characterises the majority of the data sets, and for those instances where the POF-representation was not good an *essential* objective set could still be obtained: (i) by the iterative approach of the framework and (ii) by the algorithms *de-noising* features. Moreover, NL-MVU-PCA outperforms L-PCA owing to the fact that the former applies PCA to the kernel space that allows for accurate determination of the principal components.
- In Chapter 5, the DRP based exact algorithm when applied to  $\mathcal{N}_{\mathcal{P}}$  (unnoised signal) it could accurately identify the  $\delta$ -minimal set for  $\delta = 0$  and  $k$ -minimal error for  $k = I$  in case of DTLZ5( $I, M$ ) and  $k = M$  in case of non-redundant problems. The same algorithm has failed to capture the correct problem features for  $\delta$ -MOSS and  $k$ -EMOSS analysis corresponding to  $\delta > 0$  and  $k < M$ , respectively. The application of NL-MVU-PCA to the same problems revealed that NL-MVU-PCA is able to capture the expected trend and that the failure of the DRP exact algorithm could be attributed to the stringent assumptions that the method is based on. The same results are observed when both algorithms are applied to  $\mathcal{N}_{\epsilon}$  (noised signal) where the failure of DRP based exact algorithm can be also attributed to its sensitivity to *noise*.
- In Chapter 6, it was found that when the preference information of the decision support in a particular iteration was erroneous, the proposed *online* implementation that features a probabilistic retention of all objectives at a given instant is able to improve the correctness of the decision support in subsequent iterations of the framework. This

has been demonstrated in comparison with the *offline* implementation despite higher demand for computational resources.

- In Chapter 7, it was found that the proposed *dissimilarity* measure is capable of identifying on its own when the underlying MOEA stabilises which implies that: (i) for two- and three-objective problems it identifies when the underlying MOEA converges to the POF and (ii) for MaOPs it identifies when the MOEA population has stagnated away from the POF. This provides an indication to the decision support as to when it is more meaningful to apply the machine learning based objective reduction analysis.

## 8.3 Limitations and Future Work

The main limitations of this thesis and the ideas that hold promise for future research in this direction, are described below.

### 8.3.1 Limitations

The decision support offered by the proposed framework will be beneficial in those MaOPs where not all the objectives may be *essential* to describe the POF, or not all the objectives may be equally important. In the case of MaOPs where the problem-structure is inherently such that all the objectives are *essential* and equally important, the proposed framework will reveal almost identical preference-weights for all the objectives, and the offered decision support will not be useful.

Besides the fundamental limitation of the thesis in terms of its scope, some limitations also exist for instances that fall well within its scope. It has been demonstrated that in the wake of the performance of most existing MOEAs (in the case of MaOPs), preservation of the correlation-structure is a better criterion than preservation of dominance relations, towards the perceived decision support. However, it may so happen that two objectives are in partial conflict (locally) while being globally correlated. In such instances, the *de-noising*

mechanisms of the framework may interpret the partial conflict as *noise*, and one of the objectives may be eliminated on the basis of correlation. In other words, while *de-noising* helps in cases where the data is characterised by *noise*, it may lead to inaccurate inferences in cases where some local characteristics (relationships between the objectives) are in contrast to the global characteristics may be interpreted as *noise*.

### 8.3.2 Scope for Future Work

In future, efforts may be made towards overcoming the limitation of the proposed framework corresponding to the cases where the globally correlated objectives may locally be in conflict. This limitation could be countered by clustering the objective space and applying the framework to each cluster. Though it would imply additional computational cost, it promises enhanced accuracy of the decision support for problems, where not all the objectives may be *essential* or equally important. Alternatively, techniques based on “*global alignment of linear models*” [182], such as *locally linear coordination*, *manifold charting* or *coordinated factor analysis*, may also be explored, where the idea is to first compute a collection of (local) linear models, and then perform an alignment of these linear models.

For tackling problems where all the objectives may be *essential* or equally important, the limitations of the existing concept of Pareto-dominance in modelling and simulating human decision making [71], need to be addressed. Towards it, while comparing two solutions: (i) *the number of improved objectives*—the number of objectives in which one solution is better than the other, and (ii) *the size of improvements*—the significance of the margins by which a solution is better or worse in certain objectives, need to be accounted for. The favourable implications are two fold. Firstly, imposing a higher selection pressure on solutions which are better in more objectives than the others, or on solutions in which certain objectives are better by more significant margins, promises a more meaningful approximation of the POF, at first place. Secondly, the information about such solutions may enable a more rational basis for the preferences articulated by the DMs.



Efforts may also be made to further improve the performance of the algorithms that perform relatively better than their counterparts (say, the density based MOEAs) in the case of MaOPs. For instance, the main challenge faced by the decomposition based MOEA/D is that the number of weights vectors associated with each individual become impractically large, when  $M \gg 4$ . Perhaps, a more efficient set of weight vectors could be generated if the information about the different objectives having different preference-weights; *the number of improved objectives*; and *the size of improvements*, is utilised.

It can be concluded that through its proposition of both *offline* and *online* decision support, while also shedding light on the *timing* of the decision support, this thesis has contributed to the widely ongoing efforts towards addressing the challenges posed by MaOPs. In particular, this thesis has added a new dimension to the ongoing research on MCDM based MOEAs towards dealing with MaOPs. This thesis acknowledges that many interesting questions remain unaddressed, and that it would require many different approaches to compliment each other, so that real-world MaOPs could effectively be solved.

## References

- [1] E. Zitzler, M. Laumanns, and S. Bleuler, “A tutorial on evolutionary multiobjective optimization,” in *Metaheuristics for Multiobjective Optimisation* (X. Gandibleux, M. Sevaux, K. Sörensen, and V. T’kindt, eds.), vol. 535 of *Lecture Notes in Economics and Mathematical Systems*, pp. 3–37, Springer Berlin / Heidelberg, 2004.
- [2] D. Brockhoff and E. Zitzler, “Objective Reduction in Evolutionary Multiobjective Optimization: Theory and Applications,” *Evolutionary Computation*, vol. 17, no. 2, pp. 135–166, 2009.
- [3] A. López Jaimes, C. A. Coello Coello, and D. Chakraborty, “Objective Reduction using a Feature Selection Technique,” in *Proceedings of the 10th annual conference on Genetic and evolutionary computation*, GECCO ’08, (New York, NY, USA), pp. 673–680, ACM, 2008.
- [4] L. K. Saul, K. Q. Weinberger, J. H. Ham, F. Sha, and D. D. Lee, “Spectral methods for dimensionality reduction,” in *Semisupervised Learning* (O. C. B. Schoelkopf and A. Zien, eds.), (Cambridge, MA), MIT Press, 2006.
- [5] K. Q. Weinberger and L. K. Saul, “Unsupervised learning of image manifolds by semidefinite programming,” *International Journal of Computer Vision*, vol. 70, no. 1, pp. 77–90, 2006.

- 
- [6] K. Deb and D. K. Saxena, "Searching for pareto-optimal solutions through dimensionality reduction for certain large-dimensional multi-objective optimization problems," in *IEEE Congress on Evolutionary Computation*, pp. 3353–3360, 2006.
  - [7] D. Saxena and K. Deb, "Non-linear dimensionality reduction procedures for certain large-dimensional multi-objective optimization problems: Employing correntropy and a novel maximum variance unfolding," in *Evolutionary Multi-Criterion Optimization* (S. Obayashi, K. Deb, C. Poloni, T. Hiroyasu, and T. Murata, eds.), vol. 4403 of *Lecture Notes in Computer Science*, pp. 772–787, Springer Berlin / Heidelberg, 2007.
  - [8] K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*. New York, NY, USA: John Wiley & Sons, Inc., 2001.
  - [9] K. Miettinen, *Nonlinear Multiobjective Optimization*. Norwell, MA: Kluwer Academic Publishers, 1998.
  - [10] S. Gass and T. Saaty, "The computational algorithm for the parametric objective function," *Naval Research Logistics Quarterly*, vol. 2, no. 1–2, pp. 39–45, 1955.
  - [11] L. Zadeh, "Optimality and non-scalar-valued performance criteria," *IEEE Transactions on Automatic Control*, vol. 8, pp. 59–60, January 1963.
  - [12] Y. Y. Haimes, L. S. Lasdon, and D. A. Wismer, "On a bicriterion formulation of the problems of integrated system identification and system optimization," *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-1, pp. 296–297, July 1971.
  - [13] V. Chankong and Y. Y. Haimes, *Multiobjective Decision Making: Theory and Methodology*. Elsevier Science, 1983.
  - [14] H. P. Benson, "Existence of efficient solutions for vector maximization problems," *Journal of Optimization Theory and Applications*, vol. 26, pp. 569–580, 1978.

- 
- [15] R. L. Keeney and H. Raiffa, *Decisions with Multiple Objectives: Preferences and Value Trade-Offs*. Cambridge University Press, 1993.
- [16] A. Charnes, W. W. Cooper, and R. O. Ferguson, “Optimal estimation of executive compensation by linear programming,” *Management Science*, vol. 1, no. 2, pp. 138–151, 1955.
- [17] J. P. Ignizio, *Goal programming and extensions*. Lexington, MA: Lexington Books, 1976.
- [18] J. P. Ignizio, “A review of goal programming: A tool for multiobjective analysis,” *The Journal of the Operational Research Society*, vol. 29, no. 11, pp. 1109–1119, 1978.
- [19] S. M. Lee, *Goal programming for decision analysis*. Philadelphia: Auerbach, 1972.
- [20] G. V. Reklaitis, A. Ravindran, and K. M. Ragsdell, *Engineering Optimization Methods and Applications*. New York: John Wiley & Sons, 1983.
- [21] R. J. Duffin, E. L. Peterson, and C. Zener, *Geometric Programming*. John Wiley and Sons, 1967. ISBN 0-471-22370-0.
- [22] D. B. Fogel, “The advantages of evolutionary computation,” in *Biocomputing and emergent computation: Proceedings of BCEC97*, pp. 1–11, World Scientific Press, 1997.
- [23] E. L. Ulungu and J. Teghem, “Multi-objective combinatorial optimization problems: A survey,” *Journal of Multi-Criteria Decision Analysis*, vol. 3, no. 2, pp. 83–104, 1994.
- [24] J. Sendin, O. Exler, and J. Banga, “Multi-objective mixed integer strategy for the optimisation of biological networks,” *Systems Biology, IET*, vol. 4, pp. 236–248, May 2010.
- [25] V. Dua and E. N. Pistikopoulos, “An algorithm for the solution of multiparametric mixed integer linear programming problems,” *Annals of Operations Research*, vol. 99, pp. 123–139, 2000.

- 
- [26] E. Zitzler, K. Deb, and L. Thiele, “Comparison of multiobjective evolutionary algorithms: Empirical results,” *Evolutionary Computation*, vol. 8, no. 2, pp. 173–195, 2000.
- [27] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler, “Scalable Test Problems for Evolutionary Multi-Objective Optimization,” in *Evolutionary Multiobjective Optimization: Theoretical Advances and Applications* (A. Abraham, R. Jain, and R. Goldberg, eds.), pp. 105–145, Springer, 2005.
- [28] K. Deb, A. Pratap, and T. Meyarivan, “Constrained test problems for multi-objective evolutionary optimization,” in *Evolutionary Multi-Criterion Optimization* (E. Zitzler, L. Thiele, K. Deb, C. Coello Coello, and D. Corne, eds.), vol. 1993 of *Lecture Notes in Computer Science*, pp. 284–298, Springer Berlin / Heidelberg, 2001.
- [29] S. Huband, P. Hingston, L. Barone, and L. While, “A review of multiobjective test problems and a scalable test problem toolkit,” *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 5, pp. 477–506, 2006.
- [30] T. Blicke, *Theory of Evolutionary Algorithms and Application to System Synthesis*. PhD thesis, Swiss Federal Institute of Technology Zurich, 1996.
- [31] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, “A Fast Elitist Multi-Objective Genetic Algorithm: NSGA-II,” *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2000.
- [32] E. Zitzler, M. Laumanns, and L. Thiele, “Spea2: Improving the strength pareto evolutionary algorithm,” Tech. Rep. 103, Computer Engineering and Networks Laboratory (TIK), Department of Electrical Engineering, Swiss Federal Institute of Technology (ETH) Zurich, May 2001.
- [33] D. Corne, J. Knowles, and M. Oates, “The pareto envelope-based selection algorithm for multiobjective optimization,” in *Parallel Problem Solving from Nature – PPSN*

- VI, vol. 1917 of *Lecture Notes in Computer Science*, pp. 839–848, Springer Berlin / Heidelberg, 2000.
- [34] K. Deb, M. Mohan, and S. Mishra, “Evaluating the  $\epsilon$ -Domination Based Multi-Objective Evolutionary Algorithm for a Quick Computation of Pareto-Optimal Solutions,” *Evolutionary Computation*, vol. 13, no. 4, pp. 501–525, 2005.
- [35] M. Laumanns, L. Thiele, K. Deb, and E. Zitzler, “Combining convergence and diversity in evolutionary multi-objective optimization,” *Evolutionary Computation*, vol. 10, no. 3, pp. 263–282, 2002.
- [36] E. Zitzler and L. Thiele, “Multiobjective optimization using evolutionary algorithms: A comparative case study,” in *Parallel Problem Solving from Nature – PPSN V*, vol. 1498 of *Lecture Notes in Computer Science*, pp. 292–301, Springer Berlin / Heidelberg, 1998.
- [37] D. Van Veldhuizen and G. Lamont, “On measuring multiobjective evolutionary algorithm performance,” in *IEEE Congress on Evolutionary Computation*, vol. 1, pp. 204–211, 2000.
- [38] E. Zitzler and S. Künzli, “Indicator-based selection in multiobjective search,” in *Parallel Problem Solving from Nature – PPSN VIII*, vol. 3242 of *Lecture Notes in Computer Science*, pp. 832–842, Springer Berlin / Heidelberg, 2004.
- [39] M. Emmerich, N. Beume, and B. Naujoks, “An emo algorithm using the hypervolume measure as selection criterion,” in *Evolutionary Multi-Criterion Optimization* (C. Coello Coello, A. Hernández Aguirre, and E. Zitzler, eds.), vol. 3410 of *Lecture Notes in Computer Science*, pp. 62–76, Springer Berlin / Heidelberg, 2005.
- [40] E. Zitzler, D. Brockhoff, and L. Thiele, “The hypervolume indicator revisited: On the design of pareto-compliant indicators via weighted integration,” in *Evolutionary Multi-Criterion Optimization* (S. Obayashi, K. Deb, C. Poloni, T. Hiroyasu, and T. Murata,

- eds.), vol. 4403 of *Lecture Notes in Computer Science*, pp. 862–876, Springer Berlin / Heidelberg, 2007.
- [41] J. Bader and E. Zitzler, “Hype: An algorithm for fast hypervolume-based many-objective optimization,” tech. rep., Computer Engineering and Networks Laboratory, ETH Zurich, 8092 Zurich, Switzerland, November 2008.
- [42] E. Hughes, “Multiple single objective pareto sampling,” in *IEEE Congress on Evolutionary Computation*, vol. 4, pp. 2678–2684, December 2003.
- [43] E. Hughes, “Msops-ii: A general-purpose many-objective optimiser,” in *IEEE Congress on Evolutionary Computation*, pp. 3944–3951, September 2007.
- [44] Q. Zhang and H. Li, “MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition,” *IEEE Transactions on Evolutionary Computation*, vol. 11, pp. 712–731, December 2007.
- [45] J. Branke, K. Deb, K. Miettinen, and R. Slowiński, eds., *Multiobjective Optimization: Interactive and Evolutionary Approaches*. Berlin, Heidelberg: Springer-Verlag, 2008.
- [46] R. Benayoun, J. de Montgolfier, J. Tergny, and O. Laritchev, “Linear programming with multiple objective functions: Step Method (STEM),” *Mathematical Programming*, vol. 1, pp. 366–375, 1971.
- [47] A. M. Geoffrion, J. S. Dyer, and A. Feinberg, “An interactive approach for multi-criterion optimization, with an application to the operation of an academic department,” *Management Science*, vol. 19, no. 4, pp. 357–368, 1972.
- [48] S. Zionts and J. Wallenius, “An interactive programming method for solving the multiple criteria problem,” *Management Science*, vol. 22, no. 6, pp. 652–663, 1976.
- [49] T. C. Koopmans, “Activity analysis of production and allocation,” *Bulletin of the American Mathematical Society*, vol. 58, no. 3, pp. 395–396, 1952.

- 
- [50] H. W. Kuhn and A. W. Tucker, "Nonlinear programming," in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, (Berkeley, Los Angeles), pp. 481–492, University of California Press, 1951.
- [51] V. Pareto, *Manuale di economia politica*. New York Public Library: Societa Editrice, 1906.
- [52] P. L. Yu, "A class of solutions for group decision problems," *Management Science*, vol. 19, no. 8, pp. 936–946, 1973.
- [53] M. Zeleny, "Compromise programming," in *Multiple Criteria Decision Making* (J. Cochrane and M. Zeleny, eds.), pp. 262–301, Columbia: University of South Carolina Press, 1973.
- [54] A. P. Wierzbicki, "Reference point approaches," in *Multicriteria Decision Making: Advances in MCDM Models, Algorithms, Theory, and Applications* (T. Gal, T. Stewart, and T. Hanne, eds.), vol. 21 of *International Series in Operations Research & Management Science*, pp. 237–275, Boston: Springer US, 1999.
- [55] P. C. Fishburn, "Lexicographic orders, utilities and decision rules: a survey," *Management Science*, vol. 20, no. 11, pp. 1442–1471, 1974.
- [56] A. Charnes and W. W. Cooper, "Management models and industrial applications of linear programming," *Management Science*, vol. 4, no. 1, pp. 38–91, 1957.
- [57] A. P. Wierzbicki, "A mathematical basis for satisficing decision making," *Mathematical Modelling*, vol. 3, no. 5, pp. 391–405, 1982.
- [58] A. Wierzbicki, "On the completeness and constructiveness of parametric characterizations to vector optimization problems," *OR Spektrum*, vol. 8, no. 2, pp. 73–87, 1986.
- [59] K. Miettinen, F. Ruiz, and A. Wierzbicki, "Introduction to multiobjective optimization: Interactive approaches," in *Multiobjective Optimization* (J. Branke, K. Deb,



- K. Miettinen, and R. Slowinski, eds.), vol. 5252 of *Lecture Notes in Computer Science*, pp. 27–57, Springer Berlin / Heidelberg, 2008.
- [60] R. C. Purshouse and P. J. Fleming, “Evolutionary Many-Objective Optimization: An Exploratory Analysis,” in *IEEE Congress on Evolutionary Computation*, pp. 2066–2073, June 2003.
- [61] P. Fleming, R. Purshouse, and R. Lygoe, “Many-objective optimization: An engineering design perspective,” in *Evolutionary Multi-Criterion Optimization* (C. Coello Coello, A. Hernández Aguirre, and E. Zitzler, eds.), vol. 3410 of *Lecture Notes in Computer Science*, pp. 14–32, Springer Berlin / Heidelberg, 2005.
- [62] E. J. Hughes, “Evolutionary many-objective optimisation: many once or one many?,” in *IEEE Congress on Evolutionary Computation*, pp. 222–227, 2005.
- [63] H. Ishibuchi, N. Tsukamoto, and Y. Nojima, “Evolutionary many-objective optimization: A short review,” in *IEEE Congress on Evolutionary Computation*, pp. 2419–2426, June 2008.
- [64] R. E. Nisbett and T. D. Wilson, “Telling more than we can know: Verbal reports on mental processes,” *Psychological Review*, vol. 84, no. 3, pp. 231–259, 1977.
- [65] P. Slovic and S. Lichtenstein, “Comparison of bayesian and regression approaches to the study of information processing in judgement,” *Organizational Behavior and Human Performance*, vol. 6, no. 6, pp. 649–744, 1971.
- [66] G. A. Miller, “The magical number seven, plus or minus two: Some limits on our capacity for processing information,” *Psychological Review*, vol. 63, no. 2, pp. 81–97, 1956.
- [67] J. B. Tenenbaum, V. de Silva, and J. C. Langford, “A global geometric framework for nonlinear dimensionality reduction,” *Science*, vol. 290, no. 22, pp. 2319–2323, 2000.

- 
- [68] A. V. Lotov, V. A. Bushenkov, and G. K. Kamenev, *Interactive Decision Maps*. Boston, MA: Kluwer, 2004.
- [69] S. Mostaghim and H. Schmeck, “Distance based ranking in many-objective particle swarm optimization,” in *Parallel Problem Solving from Nature – PPSN X*, vol. 5199 of *Lecture Notes in Computer Science*, pp. 753–762, Springer Berlin / Heidelberg, 2008.
- [70] H. Aguirre and K. Tanaka, “Working principles, behaviour, and performance of moeas on mnk-landscapes,” *European Journal of Operational Research*, vol. 181, pp. 1670–1690, 2006. Special Issue on Evolutionary Multi-objective Optimization.
- [71] M. Farina and P. Amato, “A Fuzzy Definition of “Optimality” for Many-Criteria Optimization Problems,” *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, vol. 34, no. 3, pp. 315–326, 2004.
- [72] M. Zeleny, “Optimal Design with Multiple Criteria: DE NOVO Programming Approach,” *Engineering Costs and Production Economics*, vol. 10, no. 2, pp. 89–94, 1986.
- [73] C. M. Fonseca, L. Paquete, and M. L.-I. nez, “An Improved Dimension-Sweep Algorithm for the Hypervolume Indicator,” in *IEEE Congress on Evolutionary Computation*, (Sheraton Vancouver Wall Centre Hotel, Vancouver, BC, Canada), pp. 16–21, July 2006.
- [74] A. P. Guerreiro, C. M. Fonseca, and M. T. M. Emmerich, “A fast dimension-sweep algorithm for the hypervolume indicator in four dimensions,” in *24th Canadian Conference on Computational Geometry*, (Charlottetown, Prince Edward Island, Canada), pp. 77–82, August 2012.
- [75] L. While, “A New Analysis of the LebMeasure Algorithm for Calculating Hypervolume,” in *Evolutionary Multi-Criterion Optimization* (C. Coello Coello, A. Hernández Aguirre, and E. Zitzler, eds.), vol. 3410 of *Lecture Notes in Computer Science*, pp. 326–340, Springer Berlin / Heidelberg, 2005.

- 
- [76] L. While, P. Hingston, L. Barone, and S. Huband, "A faster algorithm for calculating hypervolume," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 1, pp. 29–38, 2006.
- [77] D. Brockhoff and E. Zitzler, "Improving hypervolume-based multiobjective evolutionary algorithms by using objective reduction methods," in *IEEE Congress on Evolutionary Computation*, pp. 2086–2093, 2007.
- [78] Y. yan Tan, Y. chang Jiao, H. Li, and X. kuan Wang, "MOEA/D + uniform design: A new version of MOEA/D for optimization problems with many objectives," *Computers & Operations Research*, 2012. Article in Press.
- [79] D. W. Corne and J. D. Knowles, "Techniques for highly multiobjective optimisation: some nondominated points are better than others," in *Proceedings of the 9th annual conference on Genetic and evolutionary computation*, pp. 773–780, 2007.
- [80] J. D. Knowles and D. W. Corne, "Approximating the nondominated front using the pareto archived evolution strategy," in *Evolutionary Computation*, vol. 8, pp. 149–172, 2000.
- [81] T. Hanne, "On the convergence of multiobjective evolutionary algorithms," *European Journal of Operational Research*, vol. 117, no. 3, pp. 553–564, 1999.
- [82] J. Branke, T. KauBler, and H. Schmeck, "Guidance in evolutionary multi-objective optimization," *Advances in Engineering Software*, vol. 32, no. 6, pp. 499–507, 2001.
- [83] J. Branke and K. Deb, "Integrating user preferences into evolutionary multi-objective optimization," in *Knowledge incorporation in evolutionary computation* (Y. Jin, ed.), pp. 461–477, Hiedelberg, Germany: Springer, 2005.

- 
- [84] K. Deb, J. Sundar, U. B. R. N, and S. Chaudhuri, “Reference point based multi-objective optimization using evolutionary algorithms,” *International Journal of Computational Intelligence Research*, vol. 2, no. 6, pp. 273–286, 2006.
- [85] L. Thiele, K. Miettinen, P. J. Korhonen, and J. Molina, “A preference-based evolutionary algorithm for multi-objective optimization,” *Evolutionary Computation*, vol. 17, no. 3, pp. 411–436, 2009.
- [86] K. Deb and A. Kumar, “Interactive evolutionary multi-objective optimization and decision-making using reference direction method,” in *Proceedings of the 9th annual conference on Genetic and evolutionary computation*, GECCO ’07, (New York, NY, USA), pp. 781–788, ACM, 2007.
- [87] K. Deb and A. Kumar, “Light beam search based multi-objective optimization using evolutionary algorithms,” in *IEEE Congress on Evolutionary Computation*, pp. 2125–2132, September 2007.
- [88] K. Deb, A. Sinha, P. Korhonen, and J. Wallenius, “An interactive evolutionary multi-objective optimization method based on progressively approximated value functions,” *IEEE Transactions on Evolutionary Computation*, vol. 14, pp. 723–739, October 2010.
- [89] A. Sinha, K. Deb, P. Korhonen, and J. Wallenius, “Progressively interactive evolutionary multi-objective optimization method using generalized polynomial value functions,” in *IEEE Congress on Evolutionary Computation*, pp. 1–8, July 2010.
- [90] P. L. Yu, “Habitual domains,” *Operations Research*, vol. 39, no. 6, pp. 869–876, 1991.
- [91] W. Heisenberg, “Über den anschulichen Inhalt der quantentheoretischen Kinematik und Mechanik,” *Zeitschrift für Physik*, vol. 43, pp. 172–198, 1927.

- 
- [92] D. Brockhoff and E. Zitzler, “On objective conflicts and objective reduction in multiple criteria optimization,” TIK Report 243, Institut für Technische Informatik und Kommunikationsnetze, ETH Zurich, February 2006.
- [93] D. Brockhoff and E. Zitzler, “Are all objectives necessary? on dimensionality reduction in evolutionary multiobjective optimization,” in *Parallel Problem Solving from Nature – PPSN IX*, vol. 4193 of *Lecture Notes in Computer Science*, pp. 533–542, Springer Berlin / Heidelberg, 2006.
- [94] I. T. Jolliffe, *Principal component analysis*. Springer series in statistics, Springer, 2002.
- [95] J.-W. Xu, P. Pokharel, A. Paiva, and J. Principe, “Nonlinear component analysis based on correntropy,” in *Neural Networks, 2006. IJCNN '06. International Joint Conference on*, pp. 1851–1855, 0-0 2006.
- [96] A. López Jaimes, C. Coello, and J. Urías Barrientos, “Online Objective Reduction to Deal with Many-Objective Problems,” in *Evolutionary Multi-Criterion Optimization* (M. Ehrgott, C. Fonseca, X. Gandibleux, J.-K. Hao, and M. Sevaux, eds.), vol. 5467 of *Lecture Notes in Computer Science*, pp. 423–437, Springer Berlin / Heidelberg, 2009.
- [97] O. Roudenko and M. Schoenauer, “A steady performance stopping criterion for Pareto-based evolutionary algorithms,” in *Proceedings of the 6th International Multi-Objective Programming and Goal Programming Conference*, 2004.
- [98] L. Martí, J. García, A. Berlanga, and J. M. Molina, “A cumulative evidential stopping criterion for multiobjective optimization evolutionary algorithms,” in *Proceedings of the 9th annual conference on Genetic and evolutionary computation*, GECCO '07, (New York, NY, USA), pp. 2835–2842, ACM, 2007.
- [99] L. Martí, J. García, A. Berlanga, and J. M. Molina, “An approach to stopping criteria for multi-objective optimization evolutionary algorithms: The MGBM criterion,” in *IEEE Congress on Evolutionary Computation*, pp. 1263–1270, May 2009.

- 
- [100] J. L. Guerrero, L. Martí, A. Berlanga, J. Garcí, and J. M. Molina, “Introducing a robust and efficient stopping criterion for MOEAs,” in *IEEE Congress on Evolutionary Computation*, pp. 1–8, July 2010.
- [101] H. Trautmann, T. Wagner, B. Naujoks, M. Preuss, and J. Mehnen, “Statistical Methods for Convergence Detection of Multi-Objective Evolutionary Algorithms,” *Evolutionary Computation*, vol. 17, no. 4, pp. 493–509, 2009.
- [102] C. A. C. Coello, G. B. Lamont, and D. A. V. Veldhuizen, *Evolutionary Algorithms for Solving Multi-Objective Problems*. Boston: Kluwer Academic Publishers, 2002.
- [103] C. A. C. Coello and G. B. Lamont, *Applications of Multi-Objective Evolutionary Algorithms*. Singapore: World Scientific, 2004.
- [104] A. Abraham, L. C. Jain, and R. Goldberg, *Evolutionary Multiobjective Optimization: Theoretical Advances and Applications*. Springer, 2005.
- [105] D. Saxena, Q. Zhang, J. a. Duro, and A. Tiwari, “Framework for many-objective test problems with both simple and complicated pareto-set shapes,” in *Evolutionary Multi-Criterion Optimization* (R. Takahashi, K. Deb, E. Wanner, and S. Greco, eds.), vol. 6576 of *Lecture Notes in Computer Science*, pp. 197–211, Springer Berlin / Heidelberg, 2011.
- [106] F. di Pierro, S.-T. Khu, and D. A. Savić, “An investigation on preference order ranking scheme for multiobjective evolutionary optimization,” *IEEE Transactions on Evolutionary Computation*, vol. 11, pp. 17–45, February 2007.
- [107] H. K. Singh, A. Isaacs, and T. Ray, “A Pareto Corner Search Evolutionary Algorithm and Dimensionality Reduction in Many-Objective Optimization Problems,” *IEEE Transactions on Evolutionary Computation*, vol. 15, pp. 539–556, August 2011.

- 
- [108] P. J. Bentley and J. P. Wakefield, “Finding acceptable solutions in the pareto-optimal range using multiobjective genetic algorithms,” in *Soft Computing in Engineering Design and Manufacturing* (C. et al., ed.), Springer Verlag, 1997.
- [109] D. Drechsler, R. Drechsler, and B. Becker, “Multi-objective optimization based on relation favour,” in *Evolutionary Multi-Criterion Optimization*, pp. 154–166, Springer Verlag, 2001.
- [110] A. Sülflow, N. Drechsler, and R. Drechsler, “Robust multi-objective optimization in high dimensional spaces,” in *Evolutionary Multi-Criterion Optimization*, vol. 4403 of *Lecture Notes in Computer Science*, pp. 715–726, Springer Berlin / Heidelberg, 2007.
- [111] F. di Pierro, S. Djordjević, Z. Kapelan, S. T. Khu, D. Savić, and G. A. Walters, “Automatic calibration of urban drainage model using a novel multi-objective genetic algorithm,” *Water Science and Technology*, vol. 52, no. 5, pp. 43–52, 2005.
- [112] I. Kokolo, H. Kita, and S. Kobayashi, “Failure of Pareto-based MOEAs: Does Non-dominated Really Mean Near to Optimal?,” in *IEEE Congress on Evolutionary Computation*, vol. 2, pp. 957–962, 2001.
- [113] H. Sato, H. Aguirre, and K. Tanaka, “Controlling dominance area of solutions and its impact on the performance of MOEAs,” in *Evolutionary Multi-Criterion Optimization* (S. Obayashi, K. Deb, C. Poloni, T. Hiroyasu, and T. Murata, eds.), vol. 4403 of *Lecture Notes in Computer Science*, pp. 5–20, Springer Berlin / Heidelberg, 2007.
- [114] H. Ishibuchi and T. Murata, “A multi-objective genetic local search algorithm and its application to flowshop scheduling,” *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 28, pp. 392–403, August 1998.
- [115] A. Jaszkievicz, “Genetic local search for multi-objective combinatorial optimization,” *European Journal of Operational Research*, vol. 137, no. 1, pp. 50–71, 2002.

- 
- [116] H. Ishibuchi, T. Doi, and Y. Nojima, “Incorporation of scalarizing fitness functions into evolutionary multiobjective optimization algorithms,” in *Parallel Problem Solving from Nature – PPSN IX*, vol. 4193 of *Lecture Notes in Computer Science*, pp. 493–502, Springer Berlin / Heidelberg, 2006.
- [117] I. Das and J. E. Dennis, “Normal-boundary intersection: A new method for generating the pareto surface in nonlinear multicriteria optimization problems,” *SIAM Journal on Optimization*, vol. 8, pp. 631–657, March 1998.
- [118] H. Ishibuchi, Y. Sakane, N. Tsukamoto, and Y. Nojima, “Simultaneous use of different scalarizing functions in MOEA/D,” in *Proceedings of the 12th annual conference on Genetic and evolutionary computation*, GECCO ’10, (New York, NY, USA), pp. 519–526, ACM, 2010.
- [119] E. Zitzler, *Evolutionary algorithms for multiobjective optimisation: methods and applications*. PhD thesis, Swiss Federal Institute of Technology, 1999.
- [120] N. Srinivas and K. Deb, “Multiobjective optimization using nondominated sorting in genetic algorithms,” *Evolutionary Computation*, vol. 2, no. 3, pp. 221–248, 1994.
- [121] D. A. V. Veldhuizen, *Multiobjective Evolutionary Algorithms: Classifications, Analyses, and New Innovations*. PhD thesis, Department of Electrical and Computer Engineering. Graduate School of Engineering. Air Force Institute of Technology, Wright-Patterson AFB, Ohio, May 1999.
- [122] T. Wagner, N. Beume, and B. Naujoks, “Pareto-, aggregation-, and indicator-based methods in many-objective optimization,” in *Evolutionary Multi-Criterion Optimization*, vol. 4403 of *Lecture Notes in Computer Science*, pp. 742–756, Springer Berlin / Heidelberg, 2007.
- [123] C. M. Fonseca and P. J. Fleming, “Genetic algorithms for multiobjective optimization: Formulation discussion and generalization,” in *Proceedings of the 5th International*



- Conference on Genetic Algorithms*, (San Francisco, CA, USA), pp. 416–423, Morgan Kaufmann Publishers Inc., 1993.
- [124] G. W. Greenwood, X. Hu, and J. G. D'Ambrosio, "Fitness functions for multiple objective optimization problems: Combining preferences with pareto rankings," in *FOGA* (R. K. Belew and M. D. Vose, eds.), pp. 437–455, Morgan Kaufmann, 1996.
- [125] M. Sakawa and K. Kato, "An interactive fuzzy satisficing method for general multi-objective 0-1 programming problems through genetic algorithms with double strings based on a reference solution," *Fuzzy Sets and Systems*, vol. 125, no. 3, pp. 289–300, 2002.
- [126] S. P. Phelps and M. Köksalan, "An interactive evolutionary metaheuristic for multi-objective combinatorial optimization," *Management Science*, vol. 49, pp. 1726–1738, December 2003.
- [127] J. K. Pekka and L. Jukka, "A visual interactive method for solving the multiple criteria problem," *European Journal of Operational Research*, vol. 24, no. 2, pp. 277–287, 1986.
- [128] A. Jaszkiewicz and R. Slowinski, "The light beam search approach - an overview of methodology applications," *European Journal of Operational Research*, vol. 113, no. 2, pp. 300–314, 1999.
- [129] T. Wagner and H. Trautmann, "Integration of preferences in hypervolume-based multiobjective evolutionary algorithms by means of desirability functions," *IEEE Transactions on Evolutionary Computation*, vol. 14, pp. 688–701, October 2010.
- [130] D. Saxena, T. Ray, K. Deb, and A. Tiwari, "Constrained many-objective optimization: A way forward," in *IEEE Congress on Evolutionary Computation*, pp. 545–552, May 2009.

- 
- [131] A. P. Wierzbicki, “The need for and possible methods of objective ranking,” in *Trends in Multiple Criteria Decision Analysis International Series in Operations Research and Management Science* (M. Ehrgott, J. R. Figueira, and S. Greco, eds.), vol. 142, pp. 37–56, Springer, 2010.
- [132] S. Beer, *Platform for Change*. New York, NY, USA: John Wiley & Sons, Inc., 1975.
- [133] M. Zeleny, *Multiple Criteria Decision Making*. New York, NY, USA: Mcgraw-Hill (Tx), 1981.
- [134] Q. Zhang, A. Zhou, and Y. Jin, “RM-MEDA: A Regularity Model-Based Multiobjective Estimation of Distribution Algorithm,” *IEEE Transactions on Evolutionary Computation*, vol. 21, pp. 41–63, February 2008.
- [135] P. Mitra, C. A. Murthy, and S. K. Pal, “Unsupervised feature selection using feature similarity,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 24, no. 3, pp. 301–312, 2002.
- [136] J. Shlens, “A tutorial on principal component analysis,” tech. rep., Center for Neural Science and Systems Neurobiology Laboratory, New York University, New York City, NY 10003-6603 and La Jolla, CA 92037, April 2009.
- [137] K. Pearson, “On lines and planes of closest fit to systems of points in space,” *Philosophical Magazine*, vol. 2, no. 6, pp. 559–572, 1901.
- [138] H. Hotelling, “Analysis of a complex of statistical variables into principal components,” *Journal of Educational Psychology*, vol. 24, no. 6, pp. 417–441, 1933.
- [139] J. B. Kruskal, “Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis,” *Psychometrika*, vol. 29, no. 1, pp. 1–27, 1964.

- 
- [140] M. A. A. Cox and T. F. Cox, *Multidimensional Scaling*, ch. Handbook of Data Visualization, pp. 315–347. Springer Handbooks of Computational Statistics, Springer Berlin / Heidelberg, 2008.
- [141] B. Schölkopf, A. Smola, and K.-R. Müller, “Nonlinear component analysis as a kernel eigenvalue problem,” *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 1998.
- [142] L. Vandenberghe and S. Boyd, “Semidefinite programming,” *SIAM Review*, vol. 38, pp. 49–95, March 1996.
- [143] M. Laumanns, L. Thiele, and E. Zitzler, “Running time analysis of evolutionary algorithms on a simplified multiobjective knapsack problem,” *Natural Computing*, vol. 3, no. 1, pp. 37–51, 2004.
- [144] L. Martí, J. García, A. Berlanga, and J. M. Molina, “A progress indicator for detecting success and failure in evolutionary multi-objective optimization,” in *IEEE Congress on Evolutionary Computation*, pp. 1–8, July 2010.
- [145] H. Trautmann, U. Ligges, J. Mehnen, and M. Preuss, “A convergence criterion for multiobjective evolutionary algorithms based on systematic statistical testing,” in *Parallel Problem Solving from Nature – PPSN X*, vol. 5199 of *Lecture Notes in Computer Science*, pp. 825–836, Springer Berlin / Heidelberg, 2008.
- [146] T. Wagner, H. Trautmann, and B. Naujoks, “OCD: Online Convergence Detection for Evolutionary Multi-Objective Algorithms Based on Statistical Testing,” in *Evolutionary Multi-Criterion Optimization* (M. Ehrgott, C. Fonseca, X. Gandibleux, J.-K. Hao, and M. Sevaux, eds.), vol. 5467 of *Lecture Notes in Computer Science*, pp. 198–215, Springer Berlin / Heidelberg, 2009.
- [147] C. M. Fonseca and P. J. Fleming, “An overview of evolutionary algorithms in multi-objective optimization,” *Evolutionary Computation*, vol. 3, pp. 1–16, March 1995.

- 
- [148] T. Goel and N. Stander, “A non-dominance-based online stopping criterion for multi-objective evolutionary algorithms,” *International Journal for Numerical Methods in Engineering*, vol. 84, no. 6, pp. 661–684, 2010.
- [149] M. Tanaka, H. Watanabe, Y. Furukawa, and T. Tanino, “Ga-based decision support system for multicriteria optimization,” in *IEEE International Conference on Systems, Man and Cybernetics*, vol. 2, pp. 1556–1561, October 1995.
- [150] A. Osyczka and S. Kundu, “A new method to solve generalized multicriteria optimization problems using the simple genetic algorithm,” *Structural and Multidisciplinary Optimization*, vol. 10, no. 2, pp. 94–99, 1995.
- [151] R. Tulshyan, K. Deb, and S. Bandaru, “Kkt proximity measure for testing convergence in smooth multi-objective optimization,” in *Proceedings of the 13th annual conference companion on Genetic and evolutionary computation*, GECCO ’11, (New York, NY, USA), pp. 93–94, ACM, 2011.
- [152] R. T. Rockafellar, *Convex Analysis*. Princeton University Press, 1996.
- [153] R. Tulshyan, R. Arora, K. Deb, and J. Dutta, “Investigating ea solutions for approximate kkt conditions in smooth problems,” in *Proceedings of the 12th annual conference on Genetic and evolutionary computation*, GECCO ’10, (New York, NY, USA), pp. 689–696, ACM, 2010.
- [154] J. Dutta, K. Deb, R. Tulshyan, and R. Arora, “Approximate kkt points and a proximity measure for termination,” *Journal of Global Optimization*, May 2012.
- [155] J. Knowles and D. Corne, “On metrics for comparing nondominated sets,” in *Congress on Evolutionary Computation*, vol. 1, pp. 711–716, May 2002.

- 
- [156] E. Zitzler, L. Thiele, M. Laumanns, C. Fonseca, and V. da Fonseca, “Performance Assessment of Multiobjective Optimizers: An Analysis and Review,” *IEEE Transactions on Evolutionary Computation*, vol. 7, pp. 117–132, April 2003.
- [157] R. W. Morrison and K. A. D. Jong, “Measurement of population diversity,” in *Selected Papers from the 5th European Conference on Artificial Evolution*, (London, UK, UK), pp. 31–41, Springer-Verlag, 2002.
- [158] M. d’Ocagne, *Coordonnées parallèles et axiales: Méthode de transformation géométrique et procédé nouveau de calcul graphique déduits de la considération des coordonnées parallèles*. Cornell University Library, 1885.
- [159] P. Jain and A. M. Agogino, “Theory of design: An optimization perspective,” *Mechanism and Machine Theory*, vol. 25, no. 3, pp. 287–303, 1990.
- [160] K. Musselman and J. Talavage, “A Tradeoff Cut Approach to Multiple Objective Optimization,” *Operations Research*, vol. 28, no. 6, pp. 1424–1435, 1980.
- [161] A. Jaszkiewicz and J. Branke, “Interactive multiobjective evolutionary algorithms,” in *Multiobjective Optimization* (J. Branke, K. Deb, K. Miettinen, and R. Slowinski, eds.), vol. 5252 of *Lecture Notes in Computer Science*, pp. 179–193, Springer Berlin / Heidelberg, 2008.
- [162] E. Hughes, “Radar waveform optimisation as a many-objective application benchmark,” in *Evolutionary Multi-Criterion Optimization* (S. Obayashi, K. Deb, C. Poloni, T. Hiroyasu, and T. Murata, eds.), vol. 4403 of *Lecture Notes in Computer Science*, pp. 700–714, Springer Berlin / Heidelberg, 2007.
- [163] E. J. Hughes, “Evan’s software page.” How published: <http://code.evanhughes.org/>.

- 
- [164] L. Gu, R. J. Yang, C. H. Tho, L. Makowski, O. Faruque, and Y. Li, “Optimization and Robustness for Crashworthiness of Side Impact,” *International Journal of Vehicle Design*, vol. 26, no. 4, pp. 348–360, 2001.
- [165] Y. T. Azene, *Work roll system optimisation using thermal analysis and genetic algorithm*. PhD thesis, School of Applied Sciences, Cranfield University, May 2011.
- [166] C. E. Shannon, “A mathematical theory of communication,” *SIGMOBILE Mobile Computing and Communications Review*, vol. 5, pp. 3–55, January 2001.
- [167] S. Kullback and R. A. Leibler, “On information and sufficiency,” *The Annals of Mathematical Statistics*, vol. 22, pp. 79–86, March 1951.
- [168] W. K. Härdle, M. Müller, S. Sperlich, and A. Werwatz, *Nonparametric and Semiparametric Models*. Springer, 2004.
- [169] D. W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons, 1992.
- [170] E. Parzen, “On estimation of a probability density function and mode,” *The Annals of Mathematical Statistics*, vol. 33, pp. 1065–1076, 1962.
- [171] M. Rosenblatt, “Remarks on some nonparametric estimates of a density function,” *The Annals of Mathematical Statistics*, vol. 27, no. 3, pp. 832–837, 1956.
- [172] T. T. Binh and U. Korn, “Mobes: A multiobjective evolution strategy for constrained optimization problems,” in *The Third International Conference on Genetic Algorithms (Mendel 97)*, pp. 176–182, 1997.
- [173] F. Kursawe, “A variant of evolution strategies for vector optimization,” in *Parallel Problem Solving from Nature – PPSN I*, (London, UK, UK), pp. 193–197, Springer-Verlag, 1991.

- 
- [174] C. Poloni, A. Giurgevich, L. Onesti, and V. Pediroda, “Hybridization of a multi-objective genetic algorithm, a neural network and a classical optimizer for a complex design problem in fluid dynamics,” *Computer Methods in Applied Mechanics and Engineering*, vol. 186, no. 2-4, pp. 403–420, 2000.
- [175] J. D. Schaffer, *Some Experiments in Machine Learning using Vector Evaluated Genetic Algorithm*. PhD thesis, Vanderbilt University, Nashville,TN, 1984.
- [176] V. R., F. C., and M. I., “Multicriteria optimization using a genetic algorithm for determining a pareto set,” *International Journal of Systems Science*, vol. 27, no. 2, pp. 255–260, 1996.
- [177] D. Saxena, J. Duro, A. Tiwari, K. Deb, and Q. Zhang, “Objective reduction in many-objective optimization: Linear and nonlinear algorithms,” *IEEE Transactions on Evolutionary Computation*, vol. 17, no. 1, pp. 77–99, 2013.
- [178] D. Freedman and P. Diaconis, “On the histogram as a density estimator: L2 theory,” *Probability Theory and Related Fields*, vol. 57, no. 4, pp. 453–476, 1981.
- [179] D. W. Scott, “On optimal and data-based histograms,” *Biometrika*, vol. 66, no. 3, pp. 605–610, 1979.
- [180] C. Stone, “An asymptotically histogram selection rule,” in *Proceedings of Second Berkeley Symposium* (J. Neyman, ed.), pp. 513–520, Berkeley: Univ. California Press, 1984.
- [181] M. Rudemo, “Empirical choice of histograms and kernel density estimators,” *Scandinavian Journal of Statistics*, vol. 9, no. 2, pp. 65–78, 1982.
- [182] L. van der Maaten, “An Introduction to Dimensionality Reduction Using Matlab,” tech. rep., Universiteit Maastricht MICC/IKAT, 2007.